Identifying historical financial crisis: Bayesian stochastic search variable selection in logistic regression

by

Chi-san Ho, B. A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

August 2009

Identifying historical financial crisis: Bayesian stochastic search variable selection in logistic regression

Approved by
Supervising Committee:
Paul Damien

Betsy Greenberg

Acknowledgements

First of all, I would like to thank my supervisor Prof. Paul Damien, for his effort and patient guidance that he provided while this report was prepared. I would especially like to thank him for his encouragement. Each time I presented him with my work, I had completely no confidence, but his warm comments and inspiring suggestions gave me the energy to carry on my research. I would also like to thank Prof. Betsy Greenberg for kindly being the reader for this report.

Next, I would like to thank Prof. Betsy Greenberg, Prof. Martha Smith and Prof. Mary Parker. It was a pleasure to work for them. As a teaching assistant, I got the chance to better understand the material and also enjoy the teaching experience.

I would also like to thank all the professors I took courses with. I especially thank Prof. Mary Parker for guiding me into the first contact with Bayesian Statistics and Prof. Robert McCulloch for his detailed explanation of this great field.

Much thanks to Hillary Patterson and Nancy Lamm for all their administrative help.

My next appreciation goes to the graduate students with whom I had the opportunity to share my live and research experiences.

Finally, I would like to thank my mother and aunt for their support in writing this report.

Abstract:

Identifying historical financial crisis: Bayesian stochastic search variable selection in logistic regression

Chi-san Ho, M.S. Stat.

The University of Texas at Austin, 2009

Supervisor: Paul Damien

This work investigates the factors that contribute to financial crises. We first study the Dow Jones index performance by grouping the daily adjusted closing value into a two-month window and finding several critical quantiles in each window. Then, we identify severe downturn in these quantiles and find that the 5^{th} quantile is the best to identify financial crises. We then matched these quantiles with historical financial crises and gave a basic explanation about them. Next, we introduced all exogenous factors that could be related to the crises. Then, we applied a rapid Bayesian variable selection technique - Stochastic Search Variable Selection (SSVS) using a Bayesian logistic regression model. Finally, we analyzed the result of SSVS, leading to the conclusion that that the dummy variable we created for disastrous hurricane, crude oil price and gold price (GOLD) should be included in the model.

Table of Contents

Identifying historical financial crisis: Bayesian stochastic search variable selection in logistic regression

Chapter 1: Introduction

The US financial market suffered severe challenges in 2008. The still ongoing financial crisis has been called the most serious one since the *Great Depression* by prestigious economists. While some experts are trying to find a remedy for this crisis, we are more interested in the cause of this rare event and in hope to find a way to predict the next one. The first thing we wanted to know is to have a clear time frame about when the crises started. As a result, we found a norm that identifies a financial crisis and then analysis the candidate factors that contributed to the crisis.

We mainly use logistic regression in this report and applied Bayesian stochastic search. Since all our data are time series, it is also important to know how many lags we should add in the model for each variable. Thus the amount of candidate variables will increase by several times.

Chapter 2: Identifying Financial Crisis

2.1 Performance of Dow Jones Index

There are many good indices which can be used for studying the economy. We picked Dow Jones Industrial Average for the following reasons. We are mainly interested in US economy and the Dow Jones Industrial Average has been established for the longest time and the data are easy to obtain. Furthermore, the component of the index is diversified enough and can well reflect the economy performance.

Our data is a daily time series from the first transaction day of 1971 to the last transaction day of 2008. In this report, we group our data in a special way which has not been used in other research to our knowledge. Our data are grouped in a two-month non-overlapping window that contains anywhere from 38 to 44 transaction days, and we treated the data in each window as a random sample from a certain distribution. For example, the first window contains data from Jan. 4, 1971 to Feb 26, 1971 and the last window (the 228th) contains data from Nov.3, 2008 to Dec.31, 2008. Thus, we can study the performance of the index by studying different quantiles in each window. The quantiles we are interested are the 5th, 25th, median, 75th and 95th.

Since the stock market grows exponentially, we take natural log of our data so that we can have a linear relationship between time and the index. Due to the fact that we have more than 200 windows, we divide the data by decades. Figure 1~4 demonstrates the distribution in each window and how they are associated with time. The red circle points out suspected financial crisis. They were at 1974, 1988, 2001, 2002 and 2008 respectively.

Figure 1:

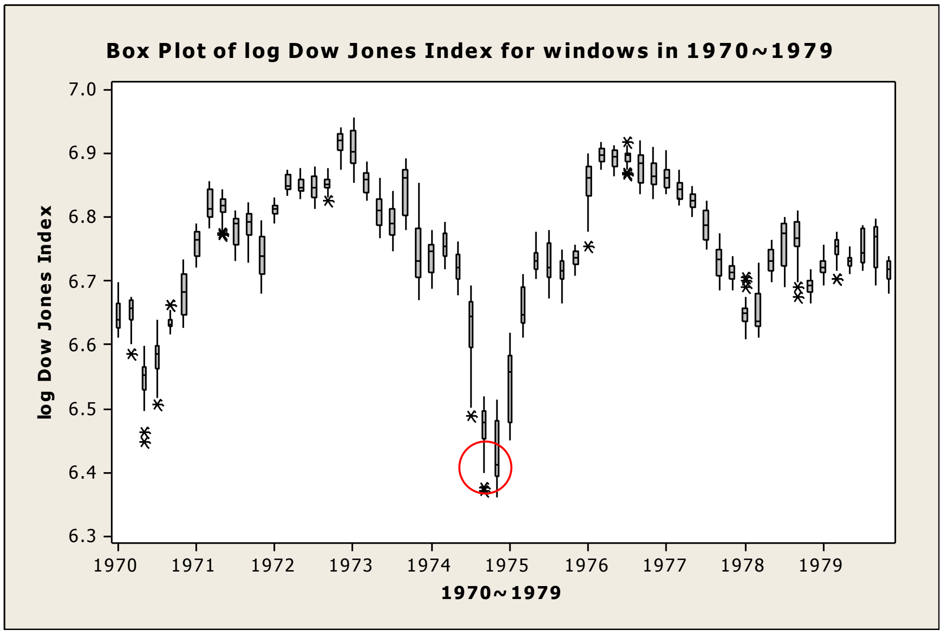


Figure 2:

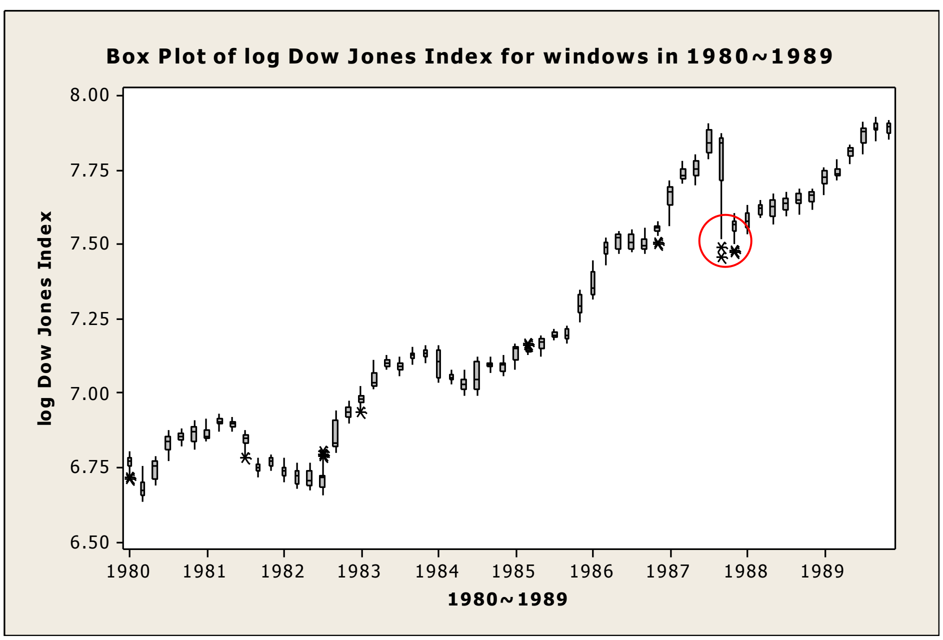


Figure 3:

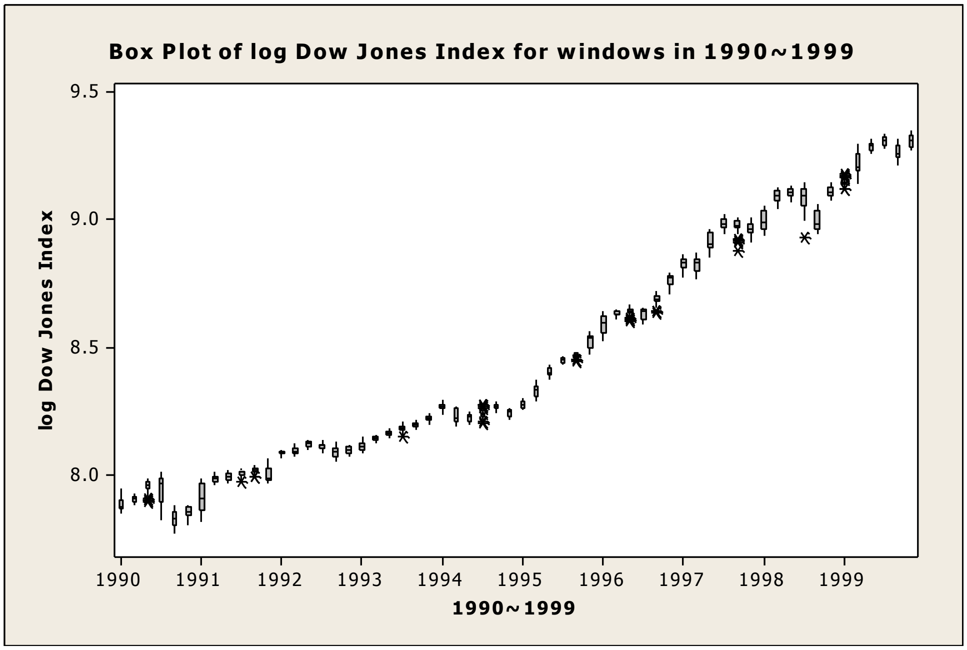
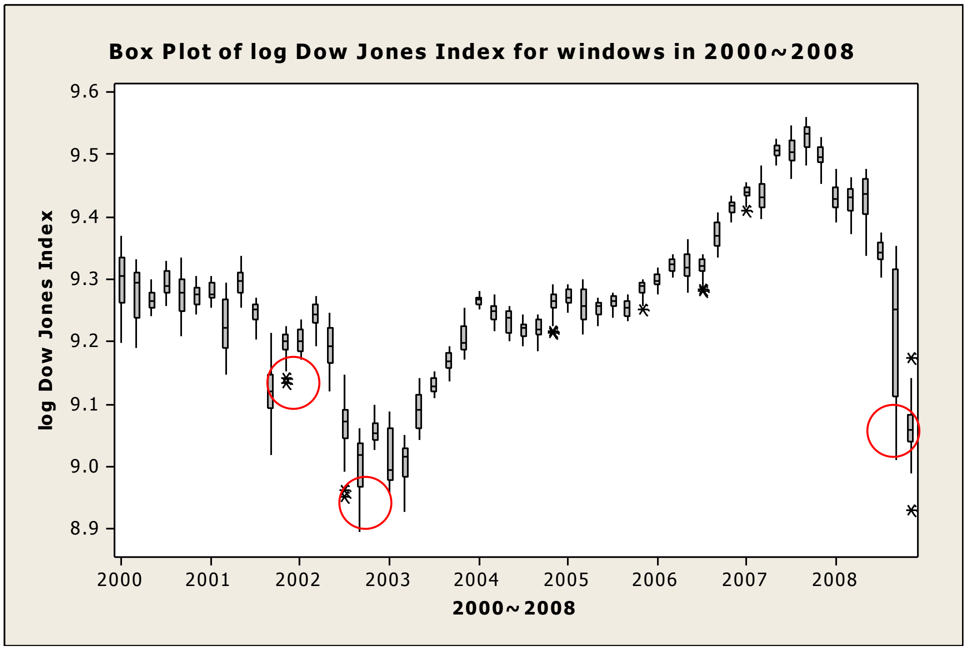


Figure 4:



Our next step is to develop a metric to identify suspected financial crisis analytically. We first examined the first order differences in those quantiles and decided to use -0.15 as a threshold to mark as a significant drop. We will verify this later. Table 1 shows us the suspected downturn for each quantile and compares the time range between the quantiles.

Table 1:

Date	Difference in the 5th quantile	Significant Drop	Difference in the first quantile	Significant Drop	Difference in the median	Significant Drop
1974 Jul. to Aug.	-17.65%	●	-10.23%		-7.67%	
1974 Sep. to Oct.	-11.15%		-14.80%		-16.81%	●
1987 Sep. to Oct.	-28.06%	●	-7.39%		-0.47%	
1987 Nov. to Dec.	-1.55%		-19.31%	●	-27.46%	●
2001 Sep. to Oct.	-16.83%	●	-14.24%		-12.96%	
2008 Sep to Oct.	-27.18%	●	-22.16%	●	-9.10%	
2008 Nov. to Dec.	-5.03%		-7.05%		-19.33%	●

Date	Difference in the third quantile	Significant Drop	Difference in the 95th quantile	Significant Drop
1974 Jul. to Aug.	-7.38%		-6.81%	●
1974 Sep. to Oct.	-17.21%	●	-16.93%	●
1987 Sep. to Oct.	-3.02%		-2.61%	
1987 Nov. to Dec.	-27.80%	●	-27.67%	●
2001 Sep. to Oct.	-11.27%		-7.11%	
2008 Sep to Oct.	-4.53%		-1.81%	
2008 Nov. to Dec.	-23.31%	●	-22.97%	●

In general, the 5^{th} quantile can detect the drop in an earlier window. It is not hard to understand the reason. When the market crashed, our window may include the data before the crash and result in better performances in other quantiles. The 5^{th} quantile is also more sensitive than the higher quantiles because it can detect the drop that other quantiles can't. As a result, we focused on the performance of the fifth quantile and use the windows detected by it to be the start of the financial crises.

Next, we would like to explore whether these sudden drops are innate. Suppose our data came from a certain distribution. In this report, we used three different inequalities to generate the upper bound of the probability of those extreme values to occur. Given we don't know anything except for the sample moments; we first applied Markov and Chebyshev inequalities.

We now show that our sample distribution is quite consistent with the normal distribution in a normal quantile plot for most of the data (more than 95%) as seen in figure 5. There are several data points (5 out of 228) that showed a deviation from the normal distribution at the lower tail. We suspect these to be outliers. Since they are a small fraction of the entire data set, we will ignore these few points and apply the normal approximation to estimate the probability of extreme events.

We set our significance level as 5%. From table 2, we can see that the normal approximation gave us the tightest estimation of the probabilities. Besides the 4 drops which we identified from our matric, we found another drop in 2002. Chebyshev inequality only gave us 1 significant drop which happened in 1987. Markov inequality was the worse among the three and failed to generate any useful information.

Figure 5: Normal Quantile Plot

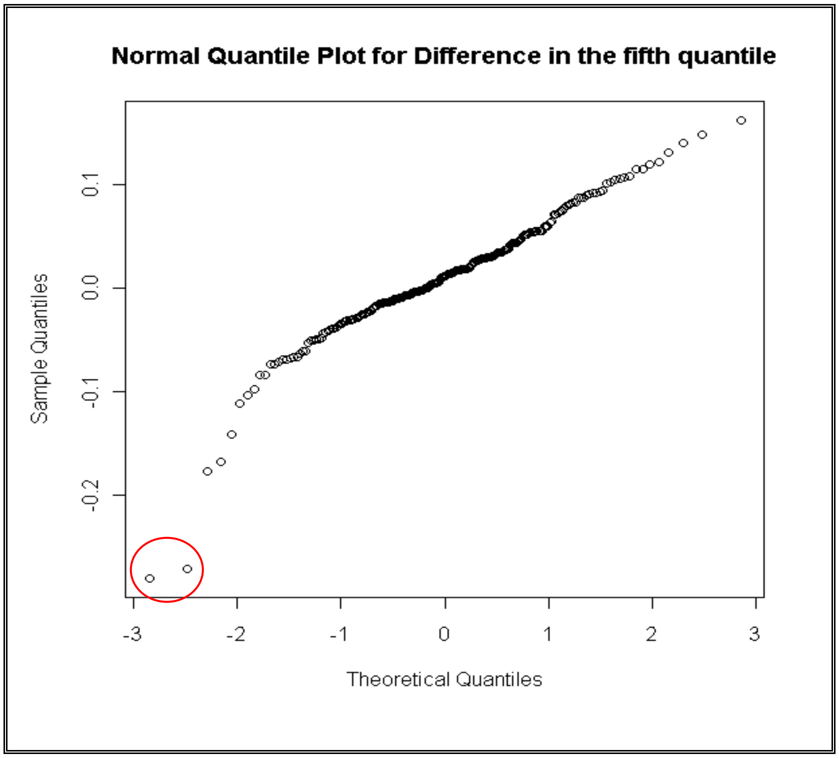


Table 2:

Date	Difference in the 5th quantile	Markov Inequality	Chebyshev Inequality	Z-score	Normal Approx.
1987 Sep. to Oct.	-28.06%	78.87%	4.93%	4.32	0.0008%
2008 Sep to Oct.	-27.18%	79.41%	5.27%	4.18	0.0015%
1974 Jul. to Aug.	-17.65%	85.84%	13.02%	2.66	0.3942%
2001 Sep. to Oct.	-16.83%	86.45%	14.41%	2.53	0.5766%
2002 Jul. to Aug.	-14.17%	88.46%	20.82%	2.10	1.7814%

We decided to use the result from normal approximation and the following research was based upon them.

2.2 Historical Financial Crisis

In this section, we would like to give a short review of these historical financial crises we identified from the 5^{th} quantile of our Dow Jones Industry Index. With the explanation from the economists and financial experts, we are able to seek our candidate variables in our model.

The crisis we identified in 1974 actually started at January 1973[1] and ended at around December 1974. This crisis affected all the major stock markets in the world. It was believed that the crash came after the collapse of the Bretton Woods system over the previous two years, with the associated "Nixon Shock" and United States dollar devaluation under the Smithsonian Agreement. The oil crisis that broke out in 1973 also compounded the market crash. Since the oil price crisis happened prior to the drop we identified, we believed it would be a good predictor in our model.

The well-known "Black Monday"[2] occurred on October 19^{th} 1987, which is the most significant one we identified. The crash began in Hong Kong and was the largest one-day percentage decline in stock market history. Although the cause of the event still continues as a debate, it is believed that the potential causes of the decline were overvaluation, illiquidity, and market psychology, especially the program trading. In our report, these factors were relatively hard to model. We could only focus on the liquidity side.

The two drops we identified in 2001 and 2002 were from a same recession period. It was called the “Early 2000s recession”[3] It was believed that the recession was caused by the move of the Federal Reserve which made successive interest increases from 2000 to 2001 and also an aftermath of the dot-com bubble. The first drop we detected in 2001 was caused by the 911 attack. The stock market was closed from Sep. 11^{th} to 17^{th} and when it was reopened, the market dropped roughly 7.1% in one day. The market rebounded after the attack but crashed once more in the second half of 2002.

The financial crisis we are experiencing now after the third quarter of 2008 is known as “Subprime mortgage crisis”[4]. It was believed to be caused by a combination of the government policy about house mortgage and the manipulation of the financial sector. Before the stock market crashed, the oil price roared to a historical high point and the gold price was also high in anticipation of the inflation. After the crisis, the housing index further plummeted and the unemployment rate rose.

Although every single crisis has its unique causes and circumstances, we would like to find the common factors. This might allow us to predict the probability of a crisis given the information provided by the covariates.

Chapter 3: SSVS Model

3.1 Forecasting Variables

In the previous chapter, we demonstrated that some of the stock market downturns are not likely to be generated by the process itself. As a result, we now would like to introduce several variables that might be able to explain the downturn. These variables are only available monthly or quarterly. In order to be consistent with our bi-month window, we modified them by either taking geometric average (for ratio or index data) or arithmetic average.

The first set of variables we are interested in are risk free investment return and the unemployment rate[5]. Since there is no perfect risk-free investment product, we used 90 Day Treasury Bill rate (IRX)[6] instead. When the economy is in a boom, the Federal Reserve will increase the interest rate to prevent inflation. On the other hand, when there are signs of recession, the Federal Reserve will keep the interest rate low. The unemployment rate is usually negatively related to the inflation rate. Also, a high unemployment rate implies poor economy. Hence, looking at the behavior of the IRX and unemployment rate only will give us an idea of health of the economy. These variables usually have reciprocal influence with the economy; however, for the sudden market crash, they might fail to provide predicting power since these data needs more time to collect and they react to the market with some lags.

We next want to add some critical ratio and price indexes. We added the crude oil price index,[7] gold price index,[8] and total credit market debt as percentage of GDP[9].

These variables affect an economy in different ways and may have some predicting power for the market crash.

Since our dependent variable is binary and only 5 observations are 1's, we lost other information provided by the previous performance of the stock market. We would like to create other variables which can provide the information we need from the performance of our quantiles. According to the theory of business cycle[10], financial crises usually came after a relatively long period of prosperity. We first created dummy variables and set them as 1 when the median increased and 0 when decreased in each window. We choose median to represent because we don't want to add too many high correlated variables in our model. Then, we added up 30 consecutive windows which represent a five-year period to be a numerical variable ranged from 0 to 30. This variable gave us a rough idea that in the previous 5 years, the stock market is going up or down generally. Our hypothesis is that if the variable has large value, it means that the economy has grown for a long period and the crisis may happen soon in the near future.

Finally, we would like to explore the relationship between the economy and the natural disasters. We searched through the historical major hurricanes[11] and created a dummy variable which is 1 when there was a hurricane during that window which caused at least a billion inflation adjusted dollar of damage.

All the variables we add up to 5 lags and the original quantiles started at the first lag. Table 3 shows the summary of all covariates and their lags.

Table 3:

Symbol	Description	Type	Frequency
IRX	US 13-Week Treasury Bill	Numerical	Monthly
UMP	US Civilian Unemployment Rate	Numerical	Monthly
COP	Price of West Texas Intermediate Crud	Numerical	Monthly
CMGDP	US Total Credit Market Debt as Percentage of GDP	Numerical	Quarterly
GOLD	Gold Price	Numerical	Monthly
HUR	Historical Disastrous Hurricane	Binary	Bi-monthly
Q5S	The Fifth Quantile of DJ Index Series	Numerical	Bi-monthly
Q50S	The Median of DJ Index Series	Numerical	Bi-monthly
SUMF	The Sum of the DIFM in 30 Consecutive Windows	Numerical	Bi-monthly
IRX1	The first lag of IRX	Numerical	Monthly
…	Etc.		

3.2 Bayesian Stochastic Search Variable Selection

Theory

The Bayesian statistical technique we used in this report is called "Stochastic Search Variable Selection" (SSVS)[12]. SSVS is for feature selection in linear regression problems; it embeds the regression set up in a hierarchical Bayesian model for identification of promising variables.

Given predictors X = [X_1, . . . ,X_k] , a dependent variable Y and a linear5 model of the form $Y = \sum_{j=1}^{k} \beta_j X_j + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$, we aim to find a selected subset of X_1, . . . ,X_k which fits the model "best". SSVS thinks promising subsets of predictors are those with higher posterior probabilities. Its goal is to avoid comparing all 2^k possible submodels by constructing an auxiliary Gibbs sequence in which the posterior distribution on the set of possible subset choices is embedded.

The idea is that if the variable X_j is very likely to be included in the model, we should have larger variance in β_j given the mean of β_j is 0. On the other hand, if the variable X_j is unlikely to be in the model, the variance of β_j should be very small and thus, the distribution of β_j is closely centered toward 0. That is, we model the uncertainty of the variable selection through a mixture normal prior. In this report, we used logistic regression[13]. So our model is slightly different from the one above. However, the principle is the same. The setting is as following:

$$Y_i \sim Bernoulli\ (P_i)\ ,\ P_i = \exp(\eta_i)/(1 + \exp(\eta_i))\ , \tag{1}$$

$$\eta_i = \sum_{j=1}^{k} \beta_j X_{i,j}\ ,\ \beta_j \sim N(0, \sigma_j) \tag{2}$$

$$\sigma_i = \gamma_i \cdot c\tau + (1-\gamma_i) \cdot \tau \tag{3}$$

We can combine (1) and (2) with a logit function:

$$\text{logit}\ ((P(Y_i = 1)) = \sum_{j=1}^{k} \beta_j X_{i,j} \tag{4}$$

We then choose c to be 1000 and τ to be 0.001 because the coefficients have the interpretation of the log-odds ratio.[14-15] This covered the range from about -0.003 to 0.003 when γ = 0 and when γ =1, it had more than 99 percent probability to cover the range from -3 to 3. Since we standardized all our data, the interpretation of the magnitude would be uniform and the range for -3 to 3 is big enough. We use the median model decision rule[16] to select all covariates where the mean of the posterior distribution of γ_j is greater than 0.5.

Computation

We implemented SSVS using a package BRugs in R-project.[17] We set the burn in value to be 2000. Because of the high autocorrelation in the Gibbs sampler, we picked 1 sample in 20 iterations. We have total 50000 iterations with 2500 sample points. We set all initial value for $\gamma_j = 1$, and β_j = 0. We set the prior distribution for the γ_j to be Bernoulli distribution with prob($\gamma_j = 1$) = 0.5.

We also had two different models. The first model (Model 1) includes only the lag covariates and the second model (Model 2) also included the concurrent covariates. The first model is predictive and the second model is for concurrent explanation.

Result

Selected covariates and the summary statistics of the posterior distributions of both models are listed in Table 4 and Table 5 and the entire results is contained in Appendix Table A and B. The MC error is less than 5% of the sample standard deviation for all parameters, which indicates that our samples are close enough to the real posterior distribution in MCMC sampling.

Table 4: Selected Covariates for Model 1

Covariates	Posterior mean	Posterior S. D	MC_error	Samples
HUR2	0.97	0.16	0.0047	2500
HUR5	0.94	0.24	0.0083	2500
HUR1	0.90	0.3	0.0101	2500
HUR4	0.90	0.3	0.0102	2500
HUR3	0.88	0.33	0.0104	2500
GOLD5	0.82	0.38	0.0139	2500
COP5	0.76	0.43	0.0156	2500
GOLD1	0.66	0.47	0.0144	2500
COP3	0.54	0.5	0.0177	2500

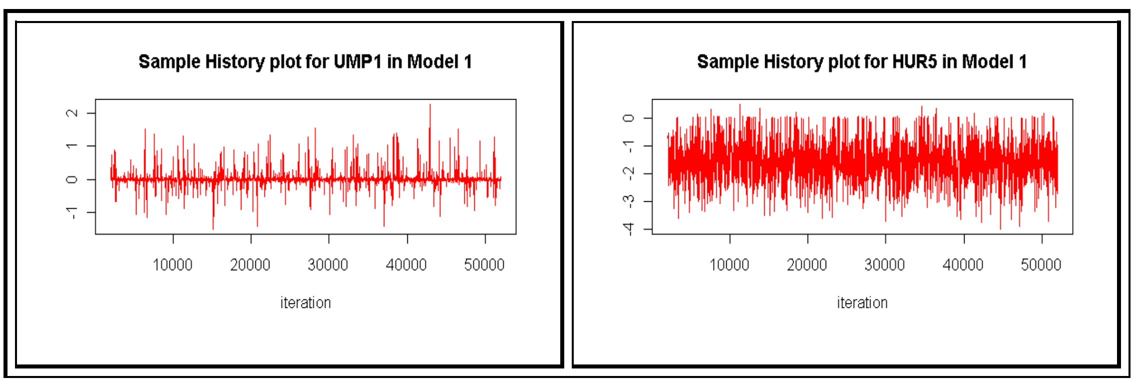
The two models yielded similar results except that Model 2 included the concurrent HUR. Also, in both models, the dummy variables (HUR and its lags) have higher probability to be chosen in SSVS procedure than other quantitative variables. In both models, they have more than 88 percent of the time to be chosen.

The history plots for the β's showed that our chains converged reasonably and gave a rough idea of the means of the posterior distributions. For example, the history plot for UMP1 appeared to be centered on 0 with relatively little deviation. On the other hand the history plot for HUR1 was denser and centered around -2. Both plots are in Table 6 and the full history plots for both models are in Appendix Table C and F.

Table 5: Selected Covariates for Model 2

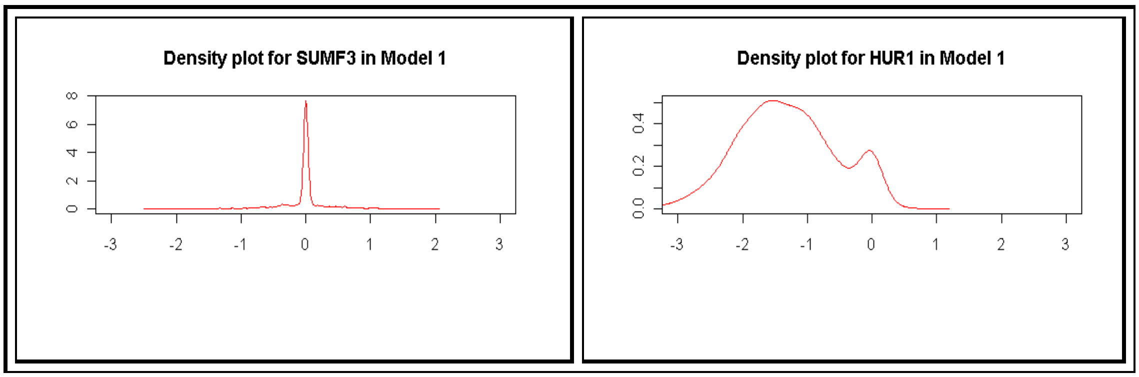
Covariates	Posterior mean	Posterior sd	MC_error	Samples
HUR2	0.922	0.27	0.0091	2500
HUR5	0.9184	0.27	0.007	2500
HUR4	0.9108	0.28	0.0084	2500
HUR3	0.9056	0.29	0.0079	2500
HUR1	0.8908	0.31	0.0109	2500
HUR	0.888	0.32	0.0097	2500
GOLD5	0.8168	0.39	0.015	2500
COP5	0.754	0.43	0.0162	2500
GOLD1	0.614	0.49	0.0163	2500
COP3	0.542	0.5	0.0189	2500

Table 6: The history plot for UMP1 and HUR1 in model 1



The density plots for the β's demonstrated the various shapes of mixture normal distributions. The density plot for SUMF1 in model 1 showed that the distribution is highly concentrated toward 0 with long tails while the density plot for HUR1 in model 1 is bimodal with two peaks, one around 0 and one around -2.5. The density plot alone gives us an idea whether this variable is chosen by SSVS. The full density plots for both models are in Appendix Table D and G.

Table 6: The history plot for UMP1 and HUR1 in model 1



The last plot we would like to examine is the autocorrelation plot for the γ's. Generally speaking, no parameters are autoregressive from the plots. The correlations decreased very fast and close 0 by lag 5 for all parameters in both models. The full autocorrelation plots for both models are in Appendix Table E and H.

Chapter 4: Conclusion and Future Research

In this report, we first identified the historical financial crises from the perspective of the stock market. We then applied SSVS to search for appropriate variables in the logistic regression model. From SSVS, we found that HUR and its lags are selected in both models. COP and GOLD are the next variables to be chosen.

We found it especially surprising that the dummy variable we created for disastrous hurricane (HUR) were selected for every single lag. At this point, we don't know how HUR affects financial crisis. We also suspected that dummy variables might have a higher chance to be selected in SSVS applied to logistic regression model with standardized exogenous variables.

This result matches our assumptions that oil prices and gold prices correlated with financial crises, and the unemployment rate and the IRX index are relatively weaker to predict financial crises. However, it is not very convincing that these three pieces of information – disastrous hurricanes, gold prices and oil prices, along can predict the occurrence of financial crises. Nevertheless, it is a good start for this extremely difficult problem. We would like to do further research on this topic and find more candidate variables.

Our future research will focus on the actual performance of the model we constructed in this report from SSVS and further explore the dummy exogenous variable issue we discovered here.

Appendix: Plots from Bayesian Analysis

Table A: Summary of the posterior distributions from model 1

Covariates	Posterior mean	Posterior sd	MC_error	Samples	Selected
HUR2	0.97	0.16	0.0047	2500	●
HUR5	0.94	0.24	0.0083	2500	●
HUR1	0.90	0.3	0.0101	2500	●
HUR4	0.90	0.3	0.0102	2500	●
HUR3	0.88	0.33	0.0104	2500	●
GOLD5	0.82	0.38	0.0139	2500	●
COP5	0.76	0.43	0.0156	2500	●
GOLD1	0.66	0.47	0.0144	2500	●
COP3	0.54	0.5	0.0177	2500	●
GOLD3	0.49	0.5	0.0133	2500	
Q5S2	0.48	0.5	0.0136	2500	
Q50S2	0.48	0.5	0.0161	2500	
GOLD2	0.47	0.5	0.0143	2500	
GOLD4	0.47	0.5	0.0137	2500	
Q5S1	0.46	0.5	0.013	2500	
Q5S3	0.46	0.5	0.0152	2500	
Q5S1	0.46	0.5	0.0142	2500	
GOLD1	0.46	0.5	0.0138	2500	
GOLD5	0.46	0.5	0.0146	2500	
GOLD2	0.46	0.5	0.0164	2500	
Q5S5	0.45	0.5	0.0154	2500	
Q50S3	0.45	0.5	0.0165	2500	
Q50S5	0.45	0.5	0.0146	2500	
COP4	0.44	0.5	0.0143	2500	
Q5S4	0.43	0.49	0.0175	2500	
Q50S4	0.43	0.5	0.013	2500	
IRX4	0.43	0.5	0.0141	2500	
GOLD3	0.43	0.49	0.0124	2500	
GOLD4	0.43	0.49	0.0135	2500	

Table A continued:

Covariates	Posterior mean	Posterior sd	MC_error	Samples	Selected
IRX3	0.39	0.49	0.0112	2500	
COP2	0.39	0.49	0.0137	2500	
SUMF3	0.36	0.48	0.0115	2500	
SUMF4	0.36	0.48	0.0126	2500	
IRX5	0.35	0.48	0.0111	2500	
COP1	0.35	0.48	0.0126	2500	
SUMF2	0.34	0.47	0.0118	2500	
SUMF5	0.34	0.47	0.0123	2500	
IRX2	0.34	0.48	0.0108	2500	
SUMF1	0.32	0.46	0.0109	2500	
IRX1	0.32	0.47	0.0111	2500	
UMP5	0.26	0.44	0.0102	2500	
UMP1	0.25	0.43	0.0126	2500	
UMP3	0.25	0.43	0.0105	2500	
UMP4	0.21	0.41	0.0104	2500	
UMP2	0.20	0.4	0.0109	2500	

Table B: Summary of the posterior distributions from model 2

Covariates	Posterior mean	Posterior sd	MC_error	Samples	Selected
HUR2	0.922	0.27	0.0091	2500	•
HUR5	0.9184	0.27	0.007	2500	•
HUR4	0.9108	0.28	0.0084	2500	•
HUR3	0.9056	0.29	0.0079	2500	•
HUR1	0.8908	0.31	0.0109	2500	•
HUR	0.888	0.32	0.0097	2500	•
GOLD5	0.8168	0.39	0.015	2500	•

COP5	0.754	0.43	0.0162	2500	●

Table B continued:

	Posterior	Posterior			
Covariates	mean	sd	MC_error	Samples	Selected
COP3	0.542	0.5	0.0189	2500	●
GOLD4	0.4816	0.5	0.016	2500	
CMGDP	0.4808	0.5	0.0162	2500	
Q5S4	0.478	0.5	0.0159	2500	
CMGDP3	0.478	0.5	0.0143	2500	
Q5S1	0.472	0.5	0.0153	2500	
Q50S3	0.4712	0.5	0.0149	2500	
CMGDP5	0.4672	0.5	0.0154	2500	
CMGDP2	0.4668	0.5	0.0141	2500	
Q5S2	0.4656	0.5	0.0148	2500	
Q50S5	0.464	0.5	0.0157	2500	
Q5S3	0.4636	0.5	0.0161	2500	
CMGDP1	0.4624	0.5	0.0152	2500	
Q5S1	0.4616	0.5	0.015	2500	
Q50S4	0.4612	0.5	0.0138	2500	
Q5S5	0.4584	0.5	0.0168	2500	
CMGDP4	0.458	0.5	0.0155	2500	
COP4	0.448	0.5	0.0127	2500	
GOLD2	0.4468	0.5	0.0154	2500	
Q50S2	0.446	0.5	0.0163	2500	
GOLD	0.4336	0.5	0.0129	2500	
IRX3	0.4252	0.49	0.0127	2500	
IRX4	0.4192	0.49	0.0146	2500	
COP2	0.392	0.49	0.0124	2500	
GOLD3	0.3888	0.49	0.0151	2500	
SUMF4	0.3836	0.49	0.0132	2500	
COP1	0.3832	0.49	0.0121	2500	
SUMF2	0.37	0.48	0.0149	2500	

SUMF3	0.3652	0.48	0.0128	2500	
SUMF1	0.3636	0.48	0.0131	2500	

Table B continued:

	Posterior	Posterior			
Covariates	mean	sd	MC_error	Samples	Selected
SUMF5	0.3504	0.48	0.012	2500	
IRX2	0.3472	0.48	0.0135	2500	
IRX5	0.3404	0.47	0.0115	2500	
IRX1	0.3316	0.47	0.011	2500	
COP	0.3104	0.46	0.0107	2500	
IRX	0.3052	0.46	0.0122	2500	
SUMF	0.3004	0.46	0.0111	2500	
UMP4	0.2924	0.45	0.0141	2500	
UMP	0.2828	0.45	0.0141	2500	
UMP2	0.2728	0.45	0.0117	2500	
UMP3	0.2624	0.44	0.0127	2500	
UMP1	0.2528	0.43	0.0107	2500	
UMP5	0.1576	0.36	0.0109	2500	

Table C: History plot for the $\beta's$ in model 1

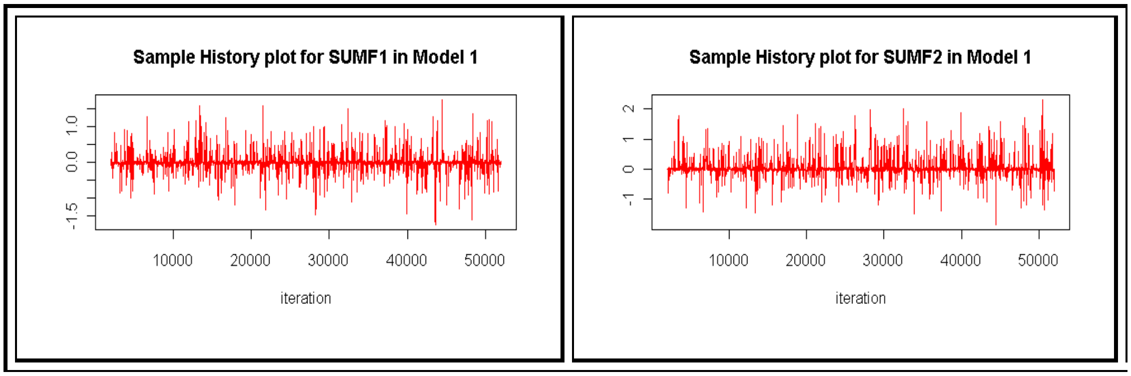
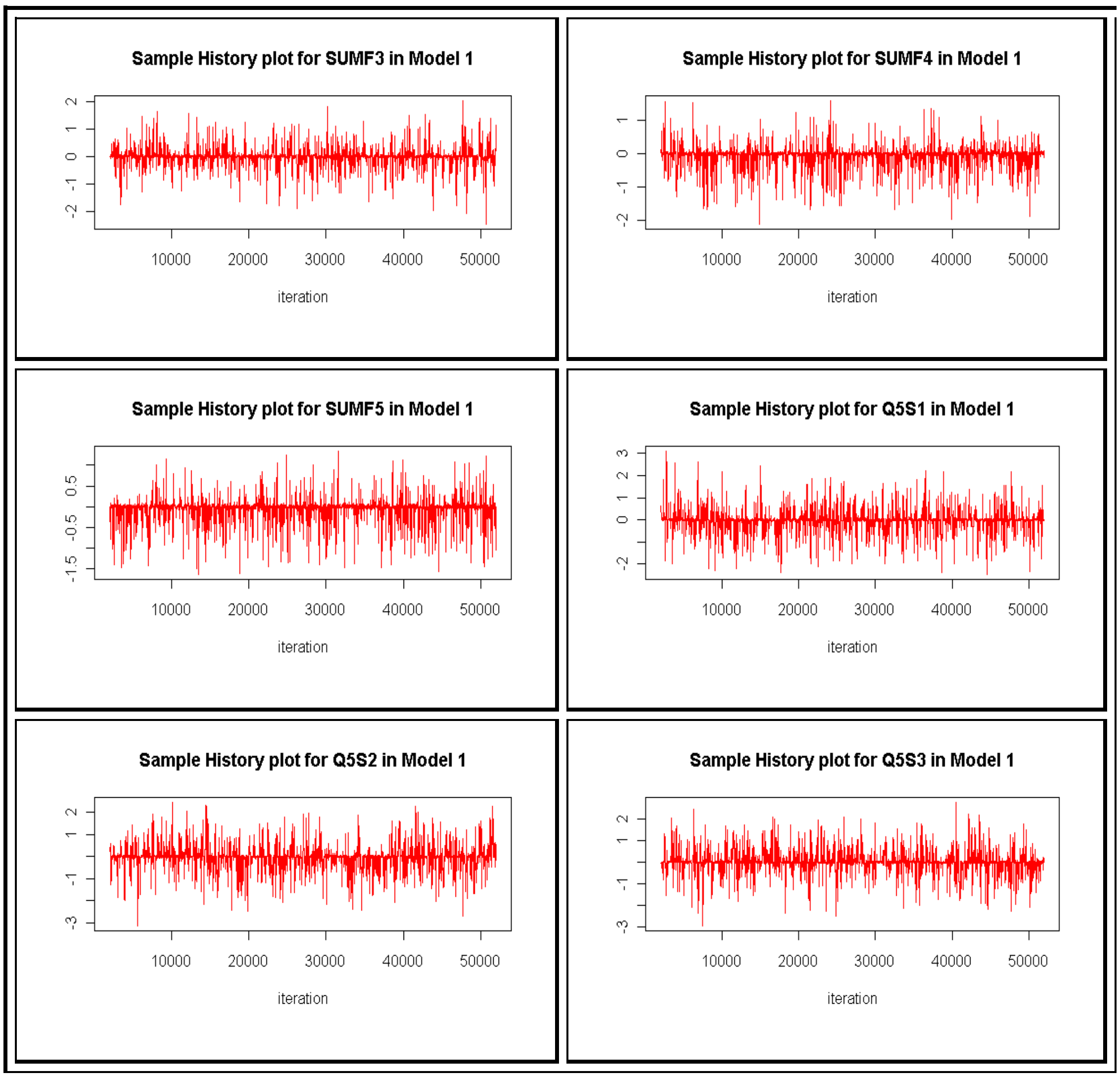
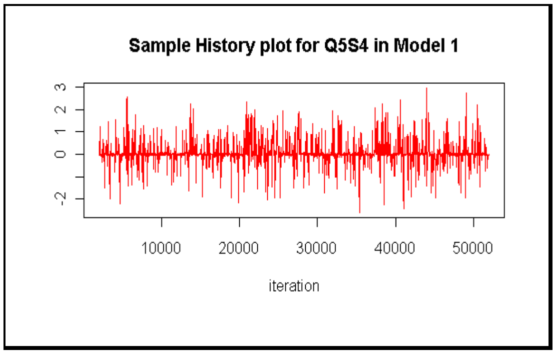


Table C continued:





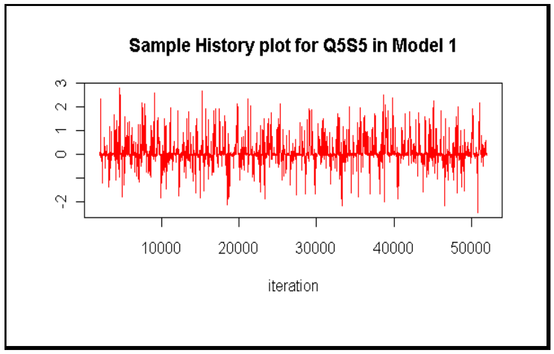
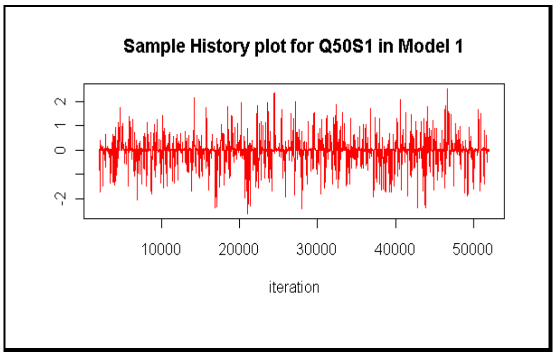
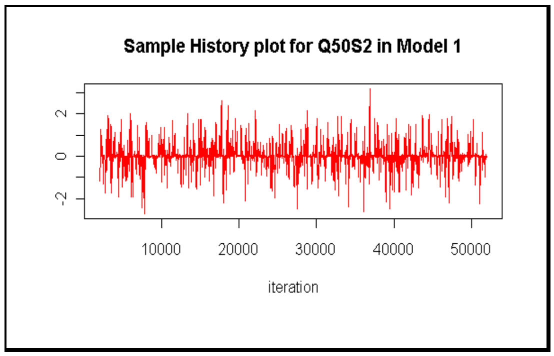
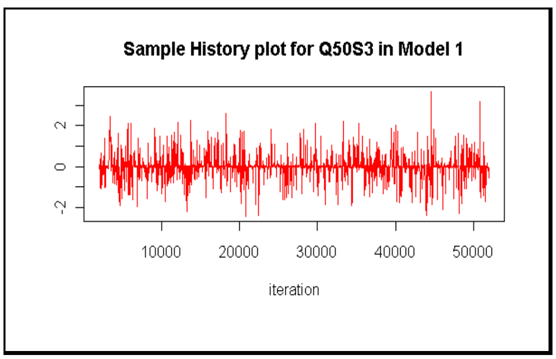
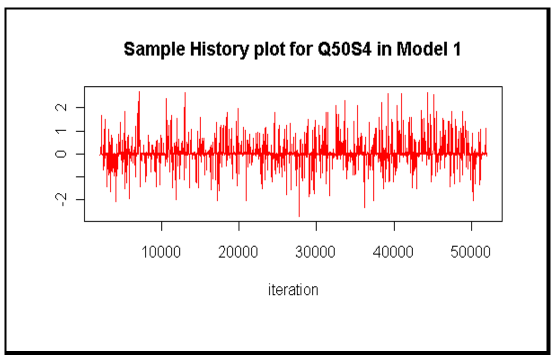


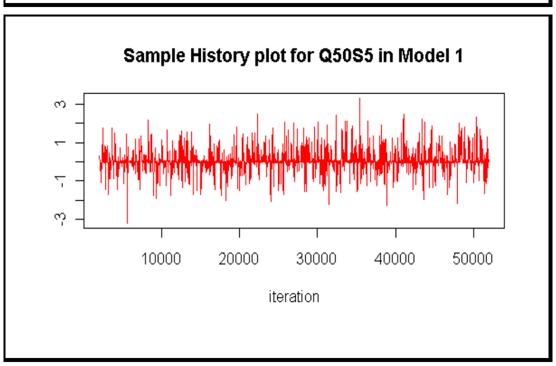
Table C continued:

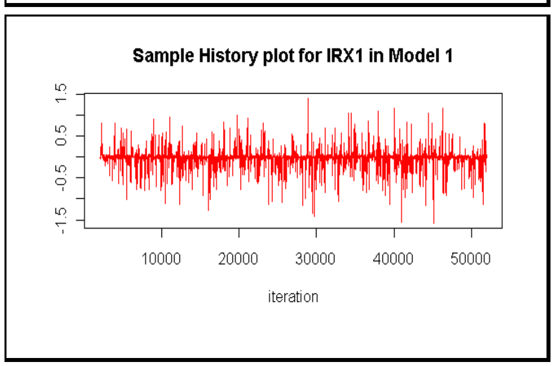


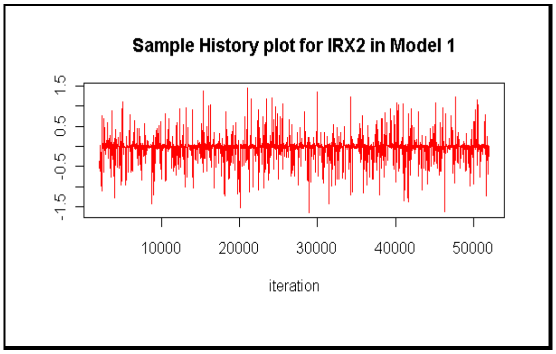












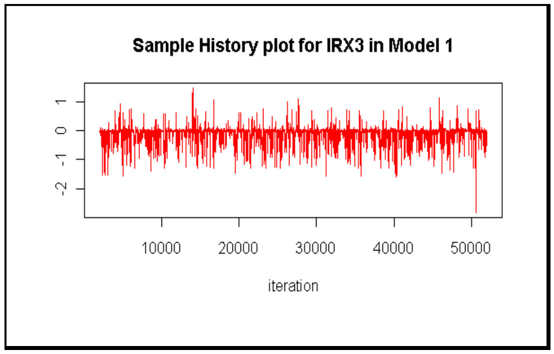
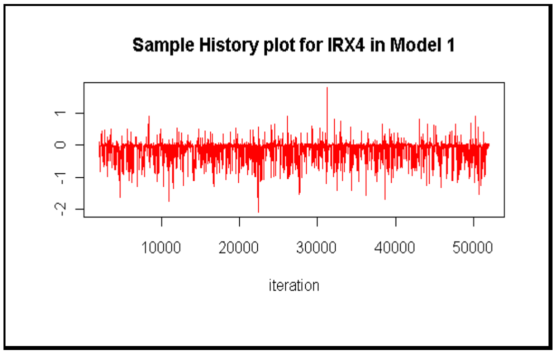
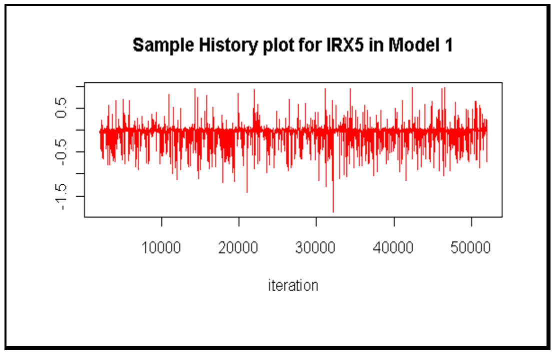
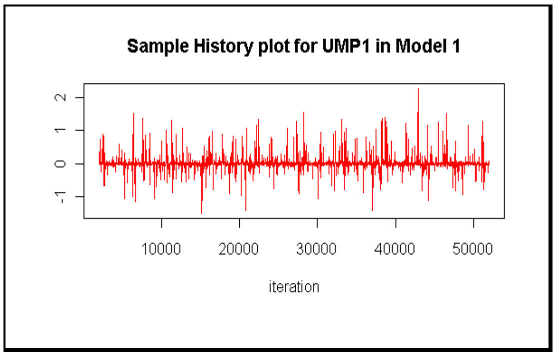
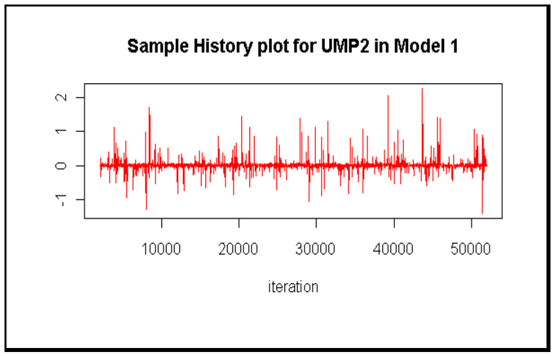


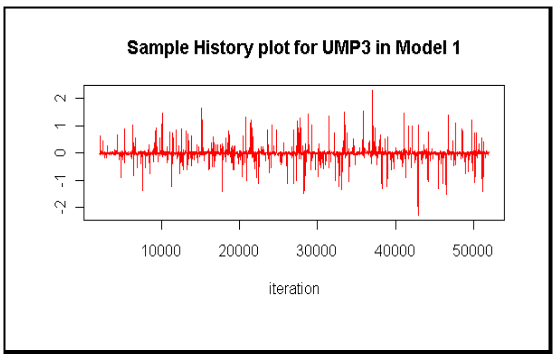
Table C continued:

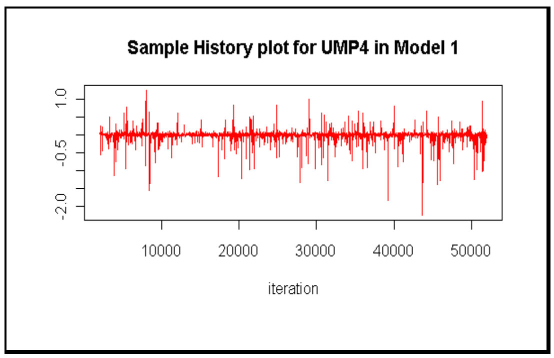


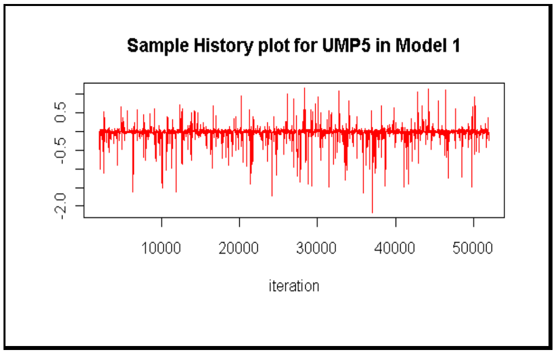












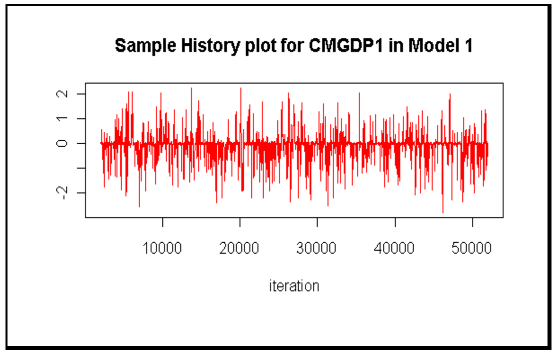
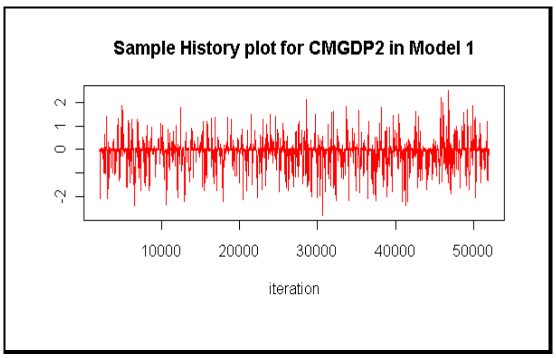
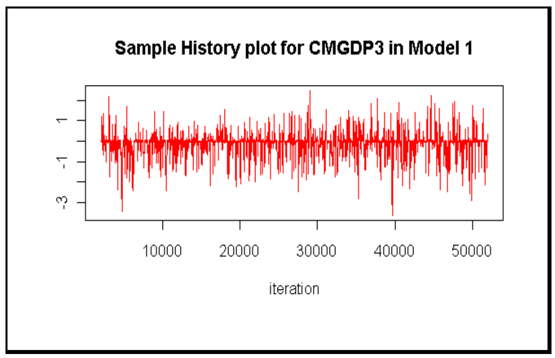
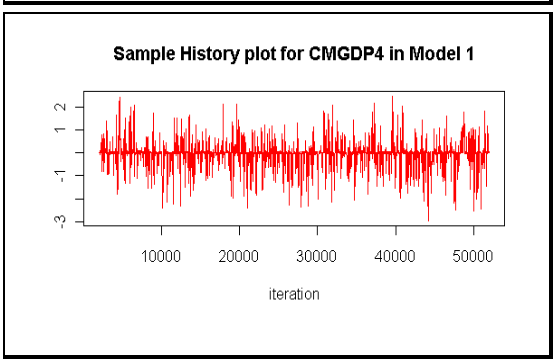
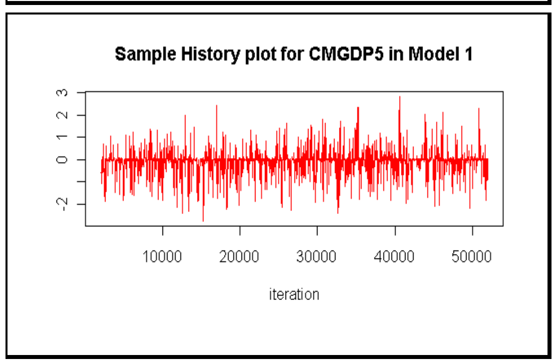


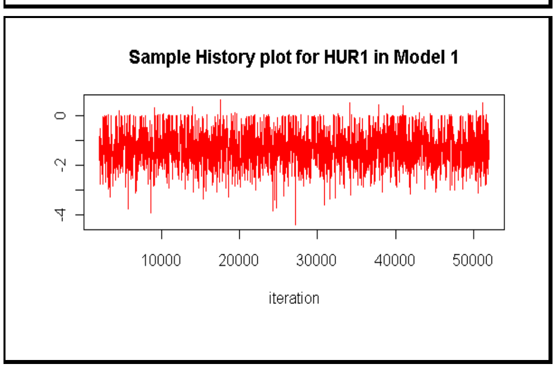
Table C continued:

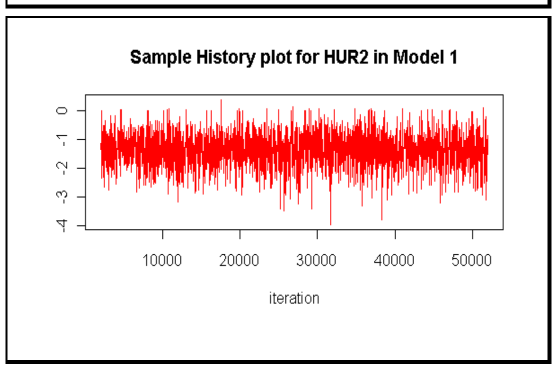












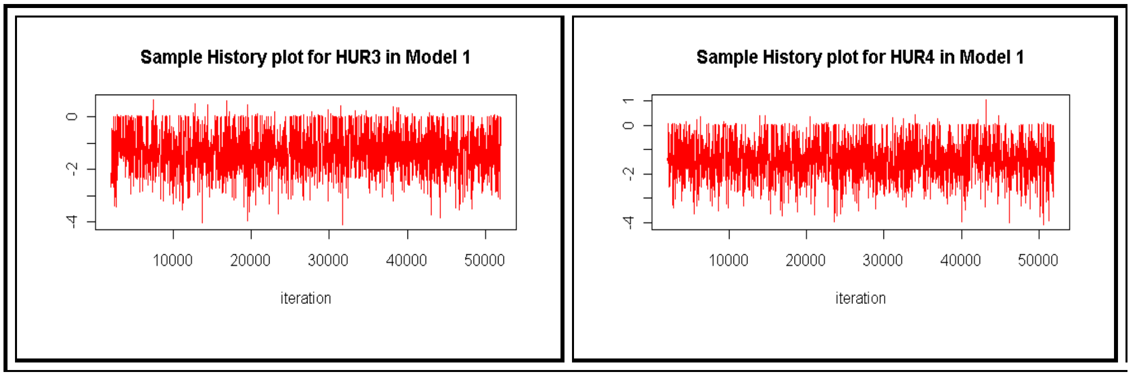
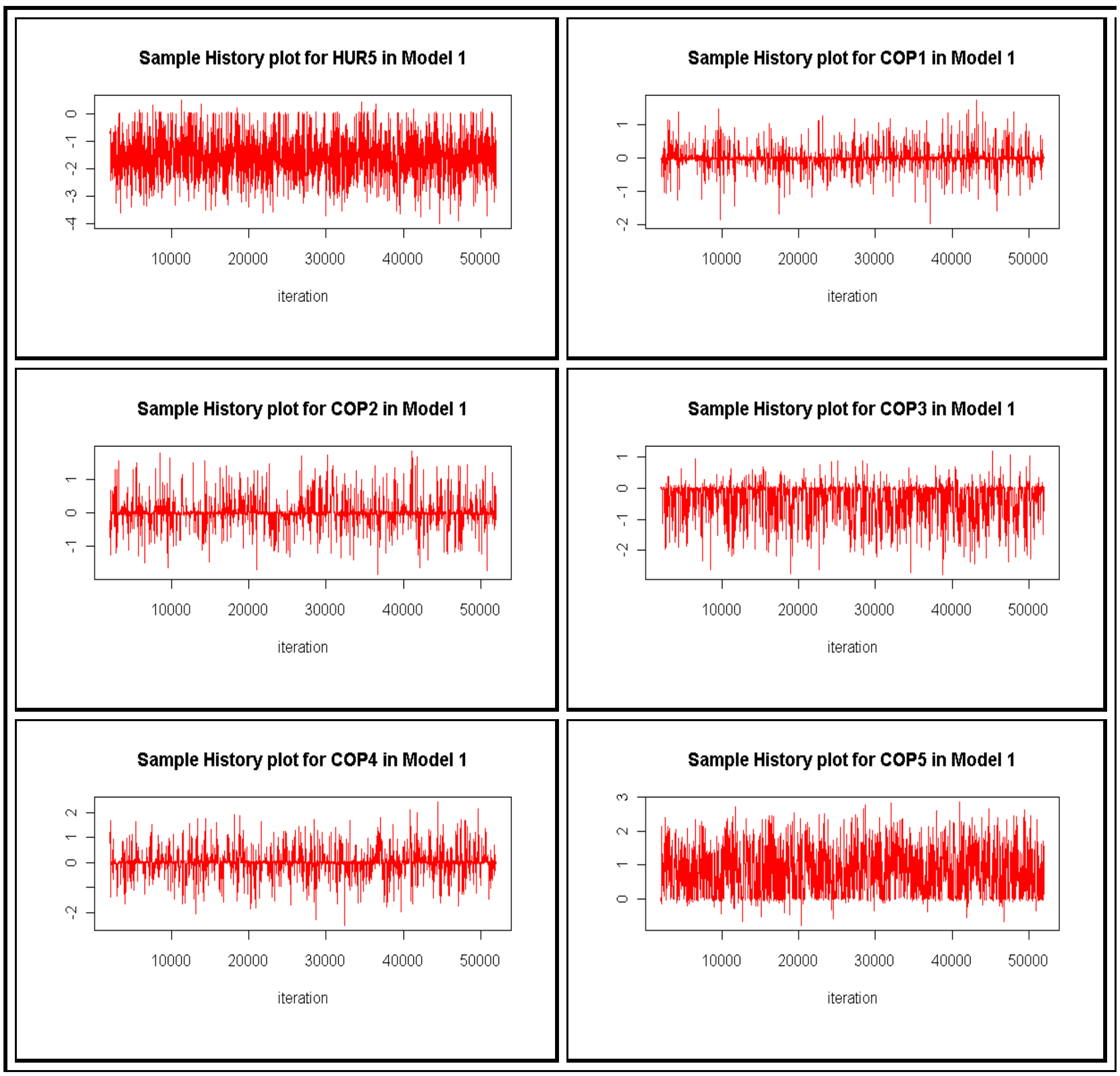
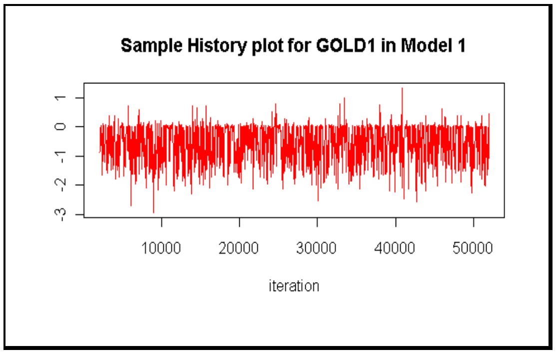


Table C continued:





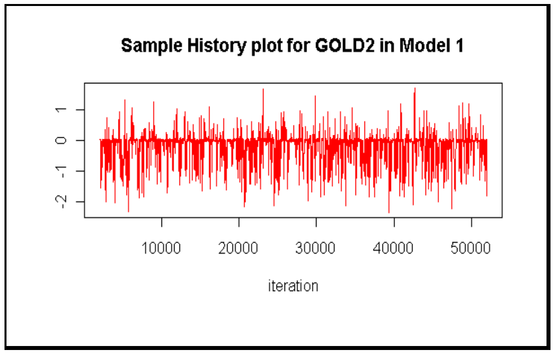
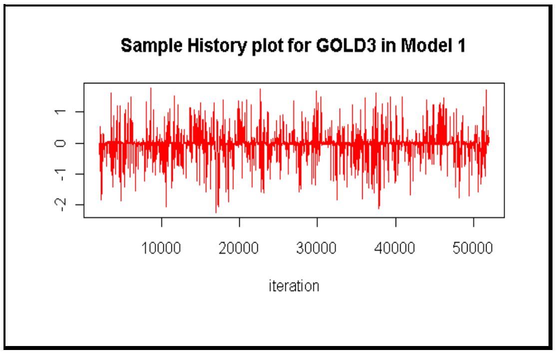
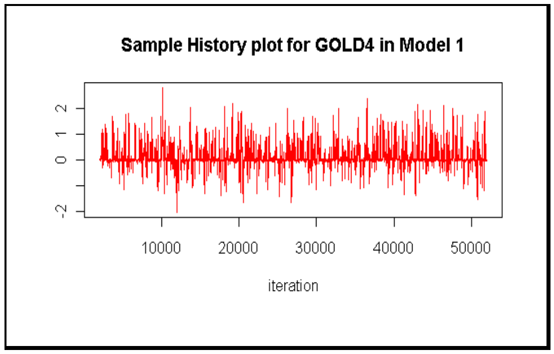


Table C continued:





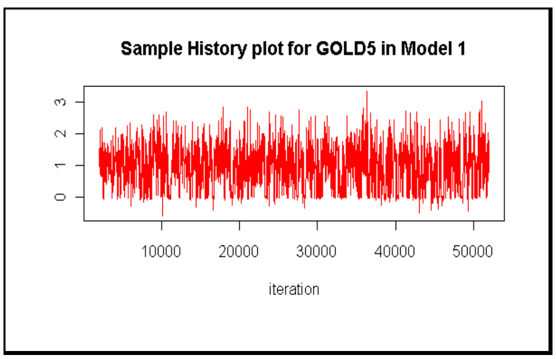
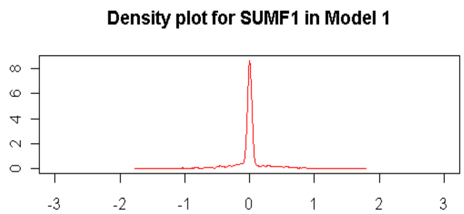


Table D: Density plot for the β's in model 1



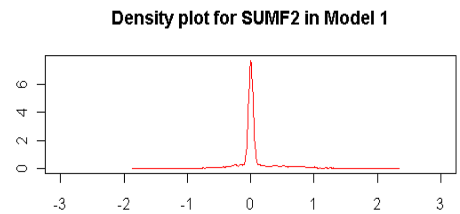
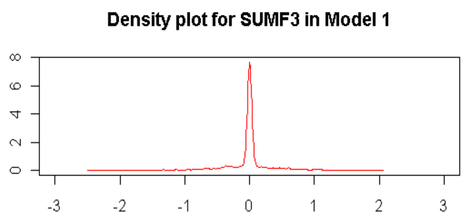
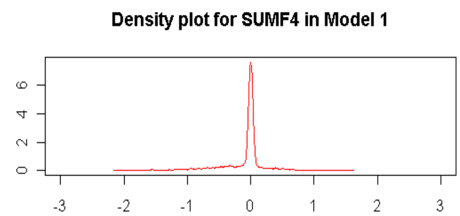
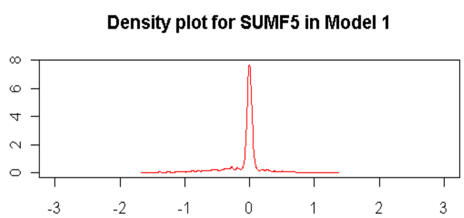
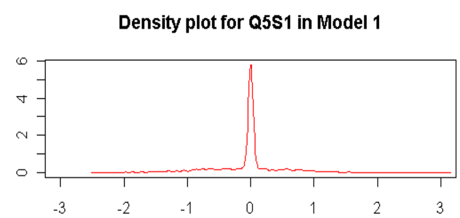


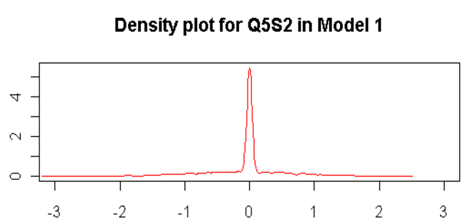
Table D continued:

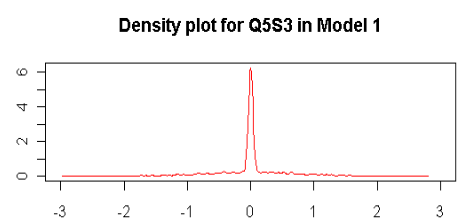


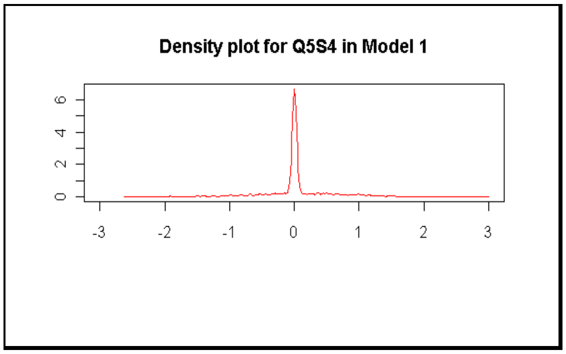












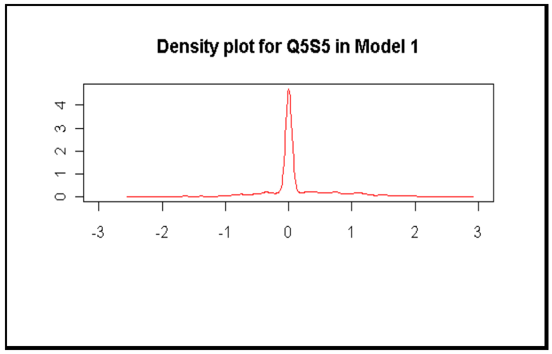
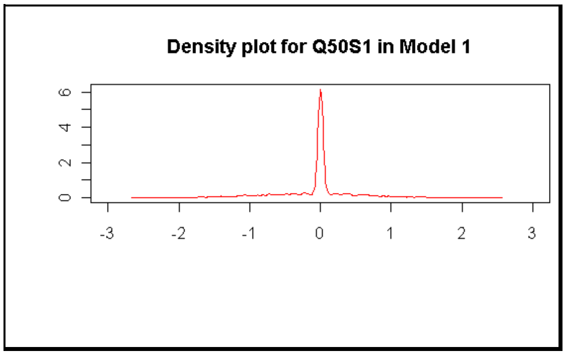
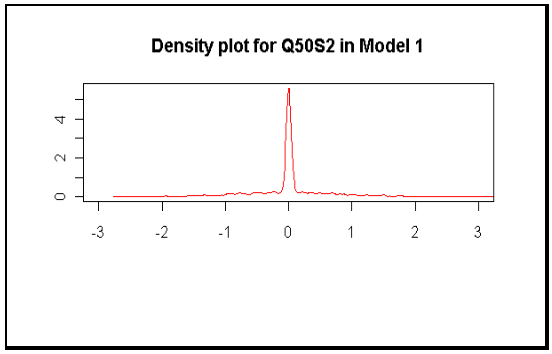
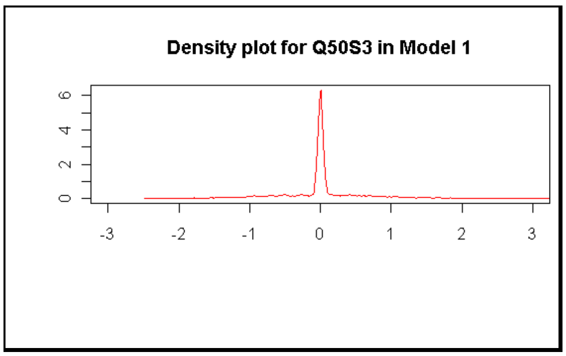
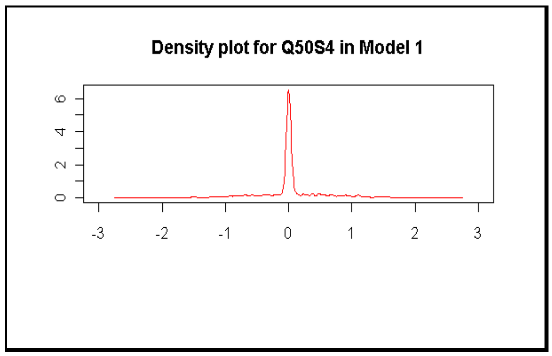


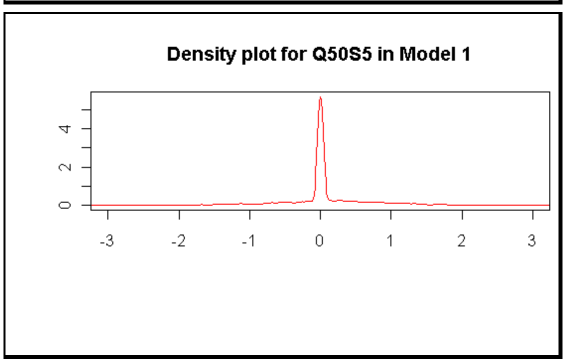
Table D continued:

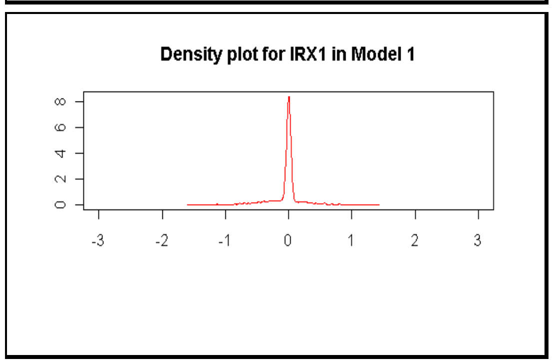












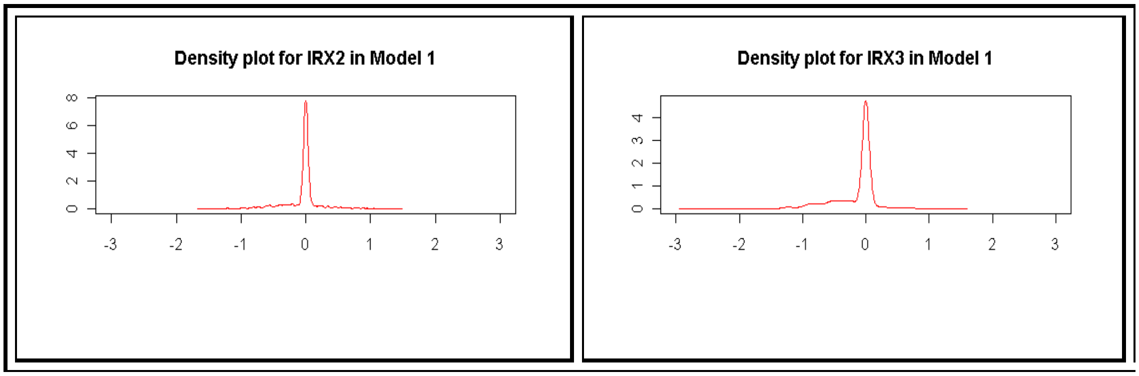
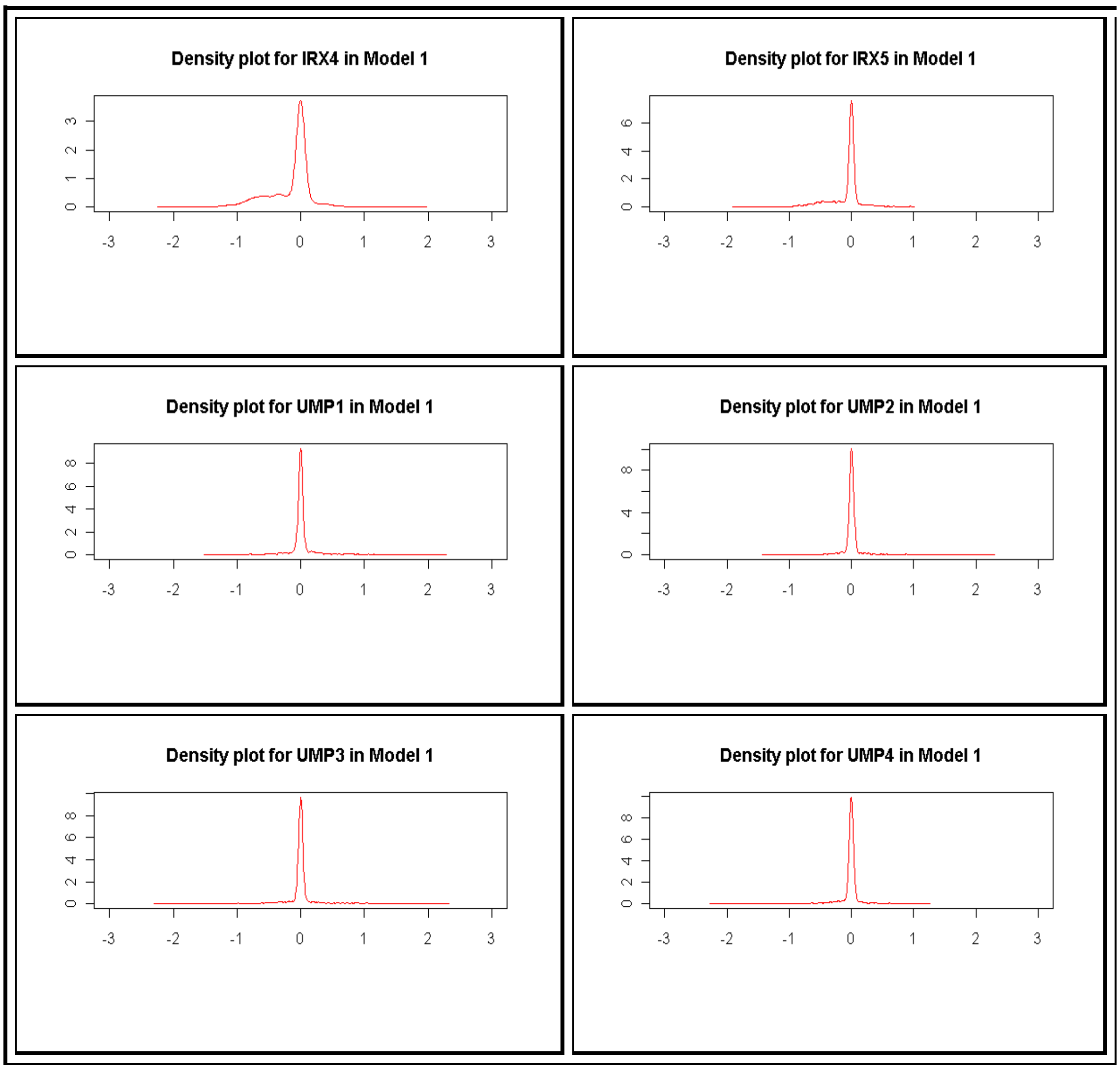
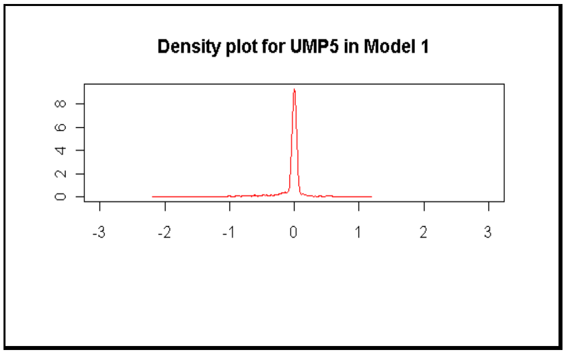


Table D continued:





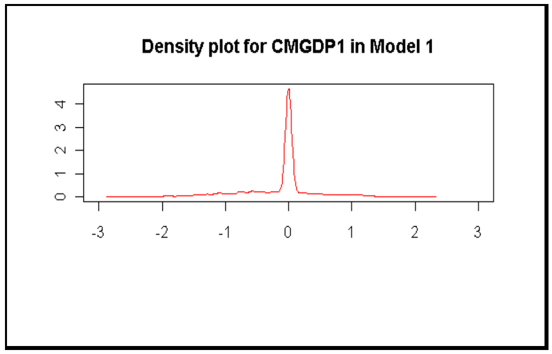
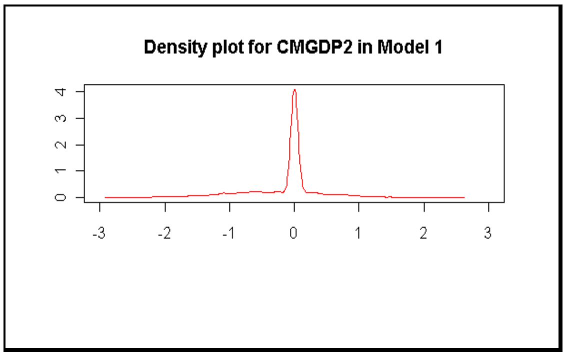
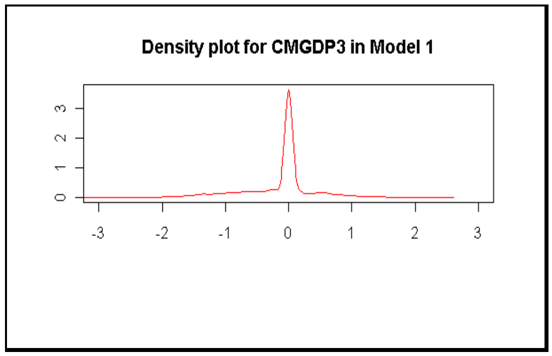
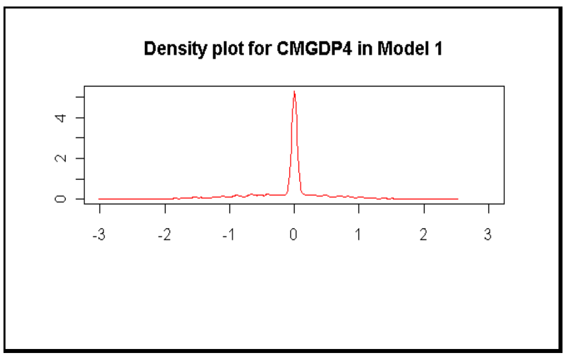
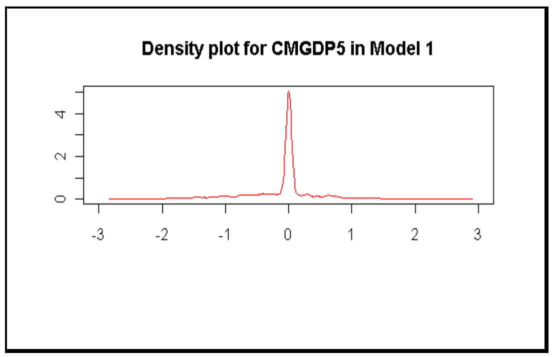


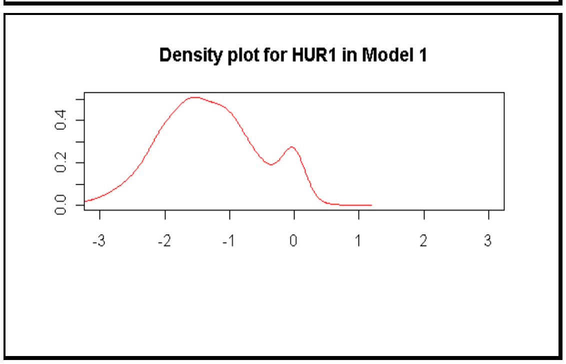
Table D continued:

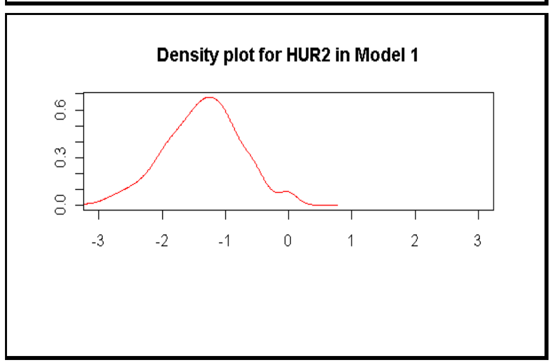


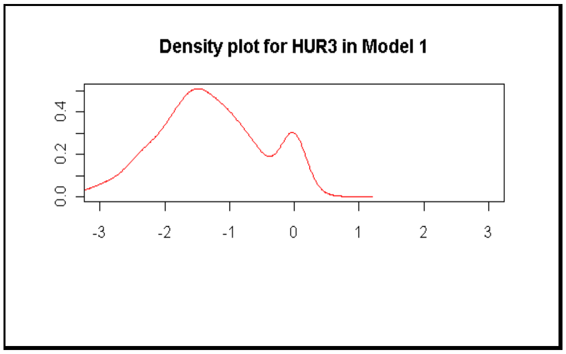












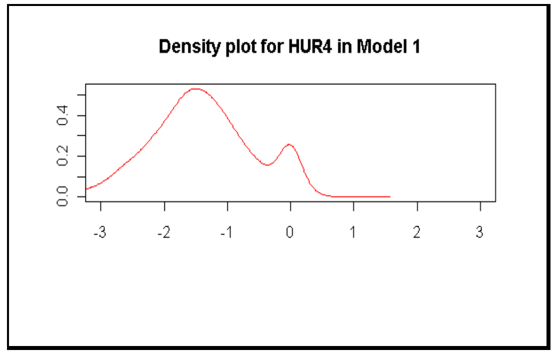
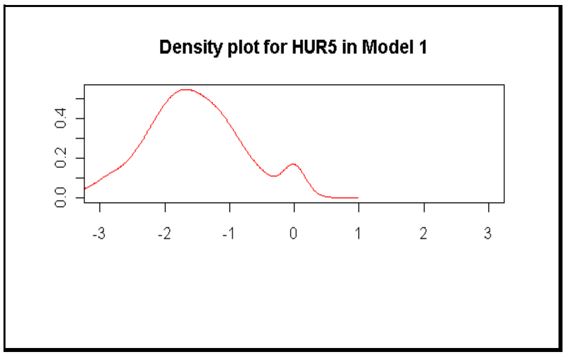
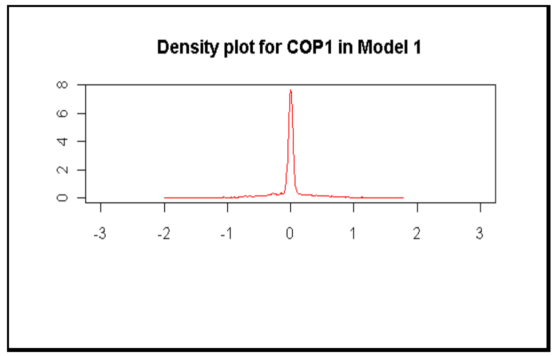
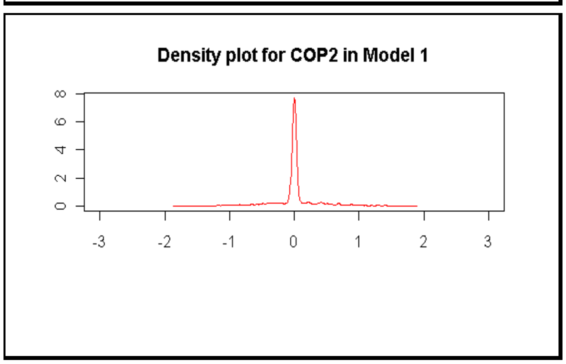
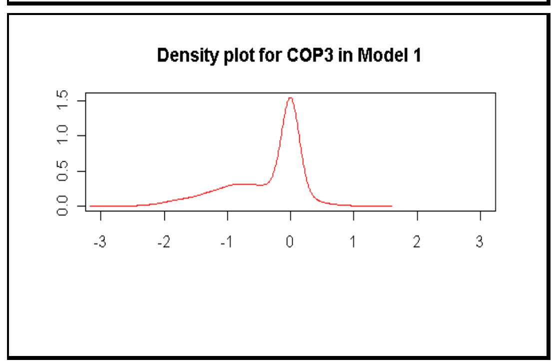


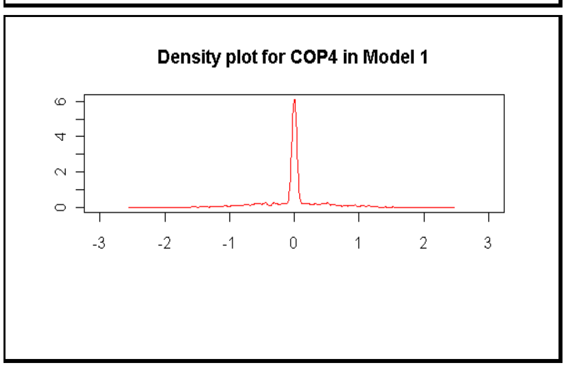
Table D continued:

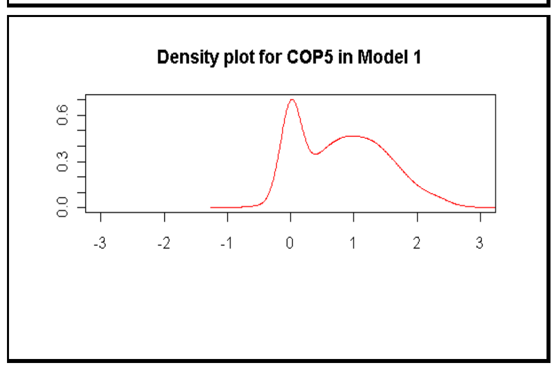












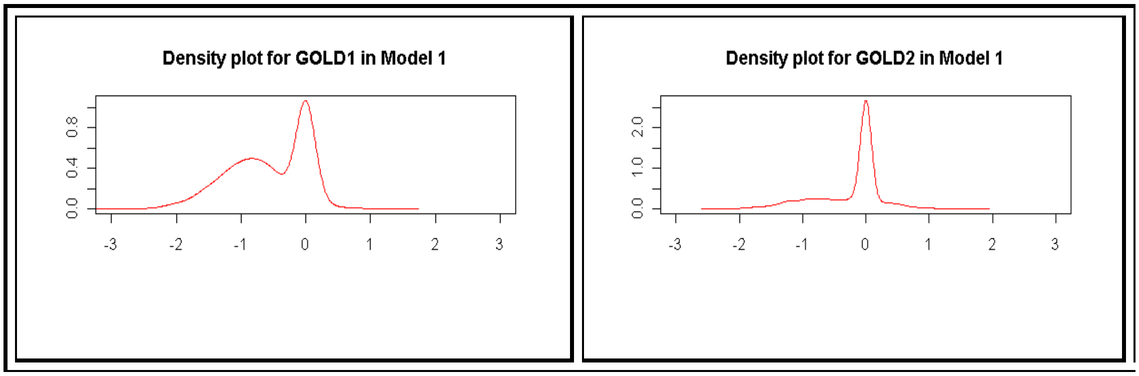


Table D continued:

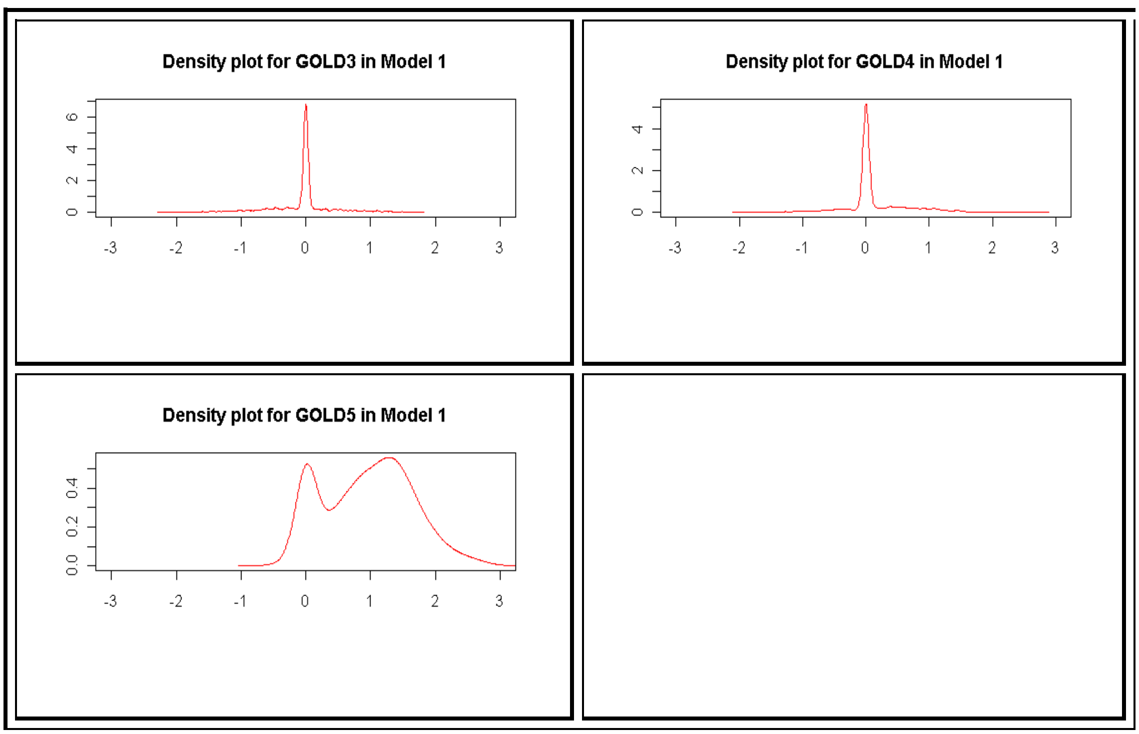
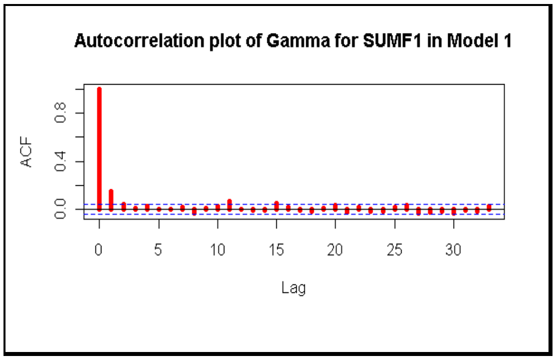


Table E: Autocorrelation plot for the γ's in model 1



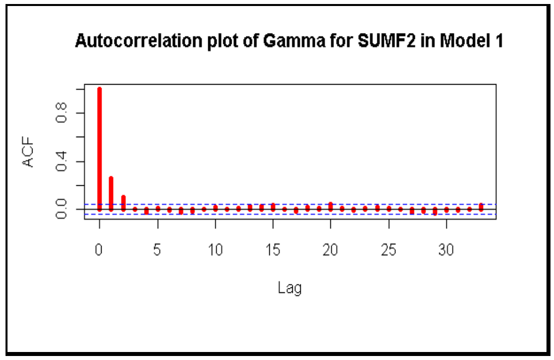
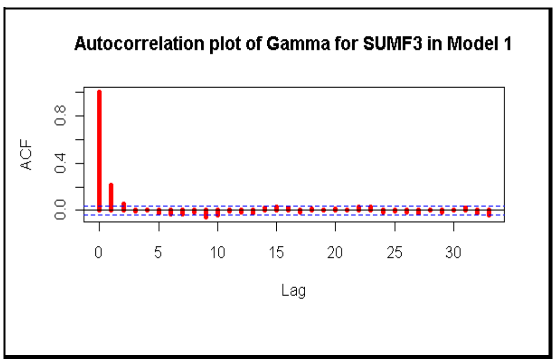
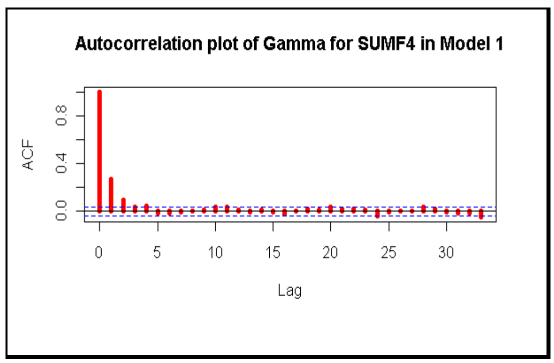
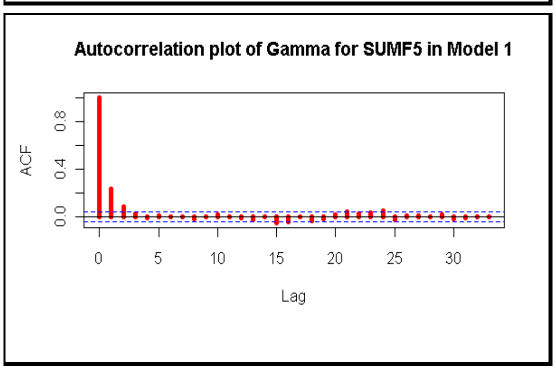
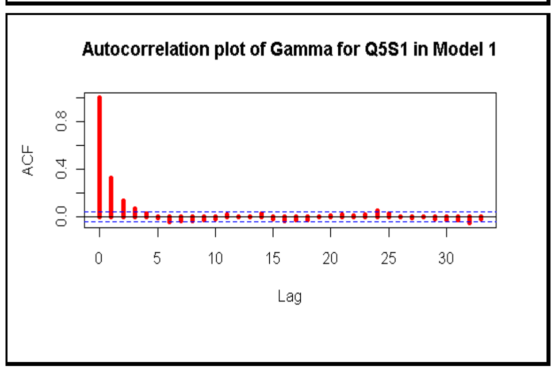


Table E continued:









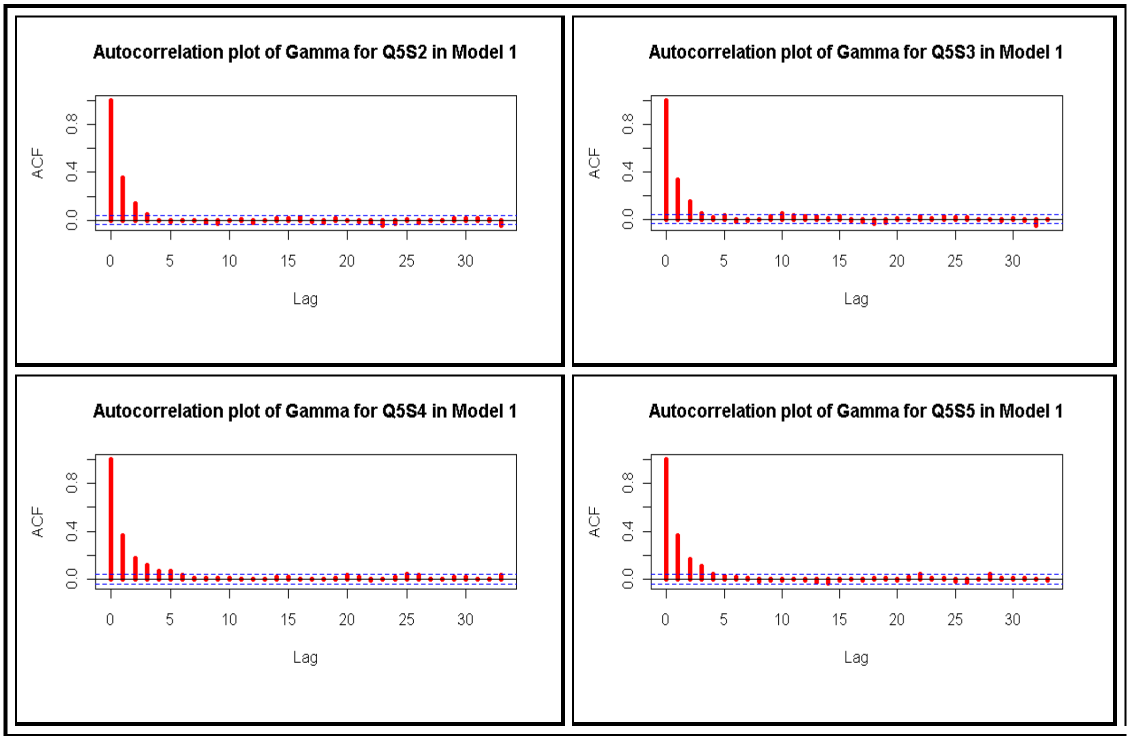
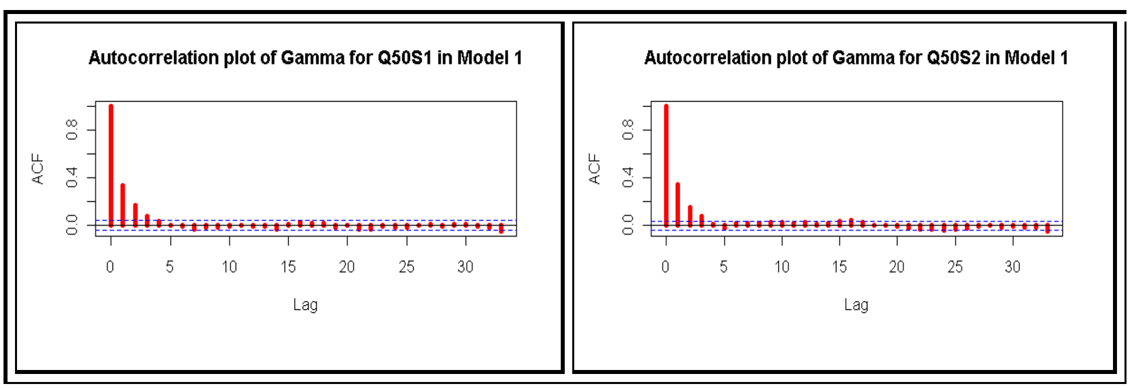


Table E continued:



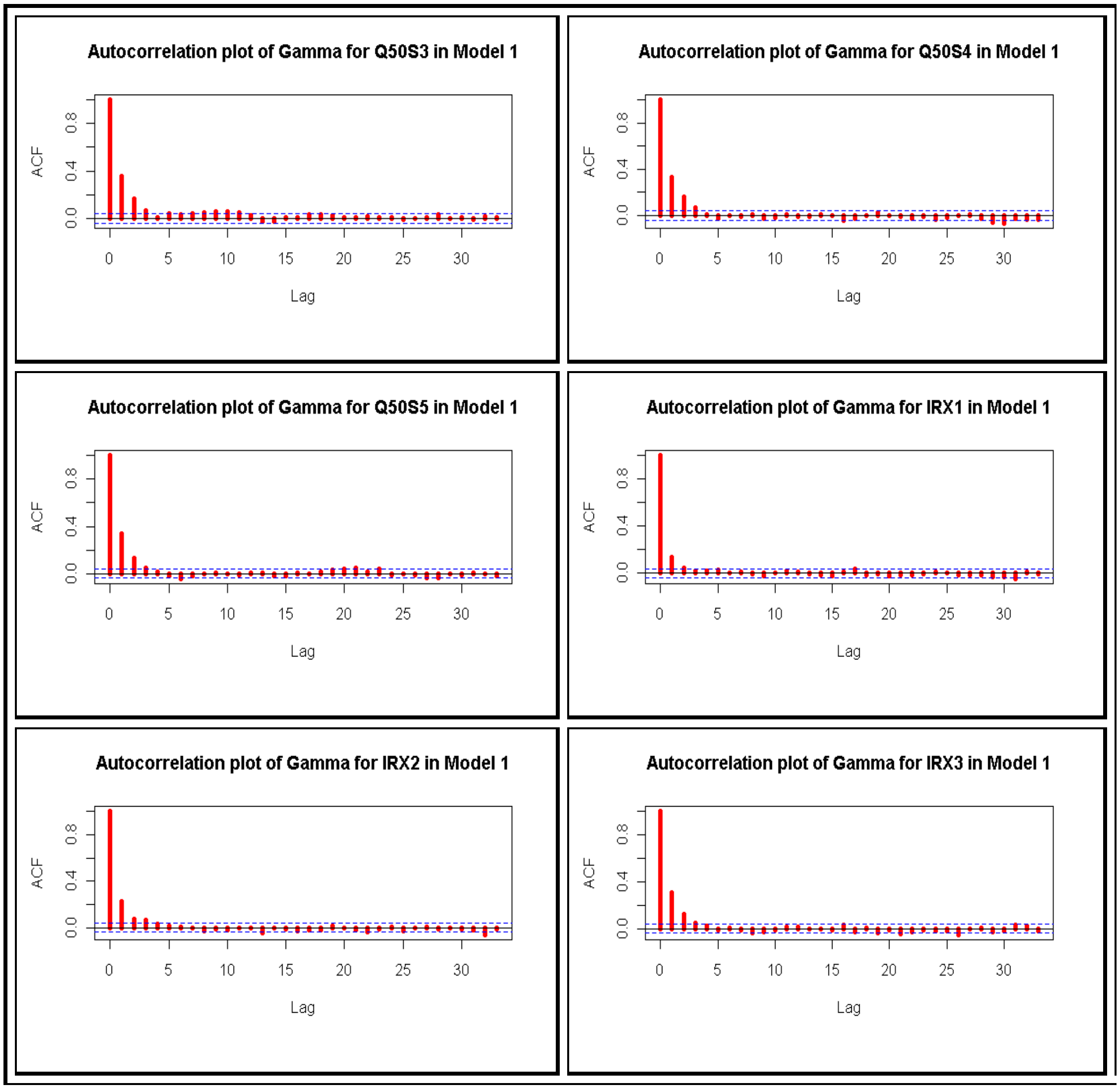


Table E continued:

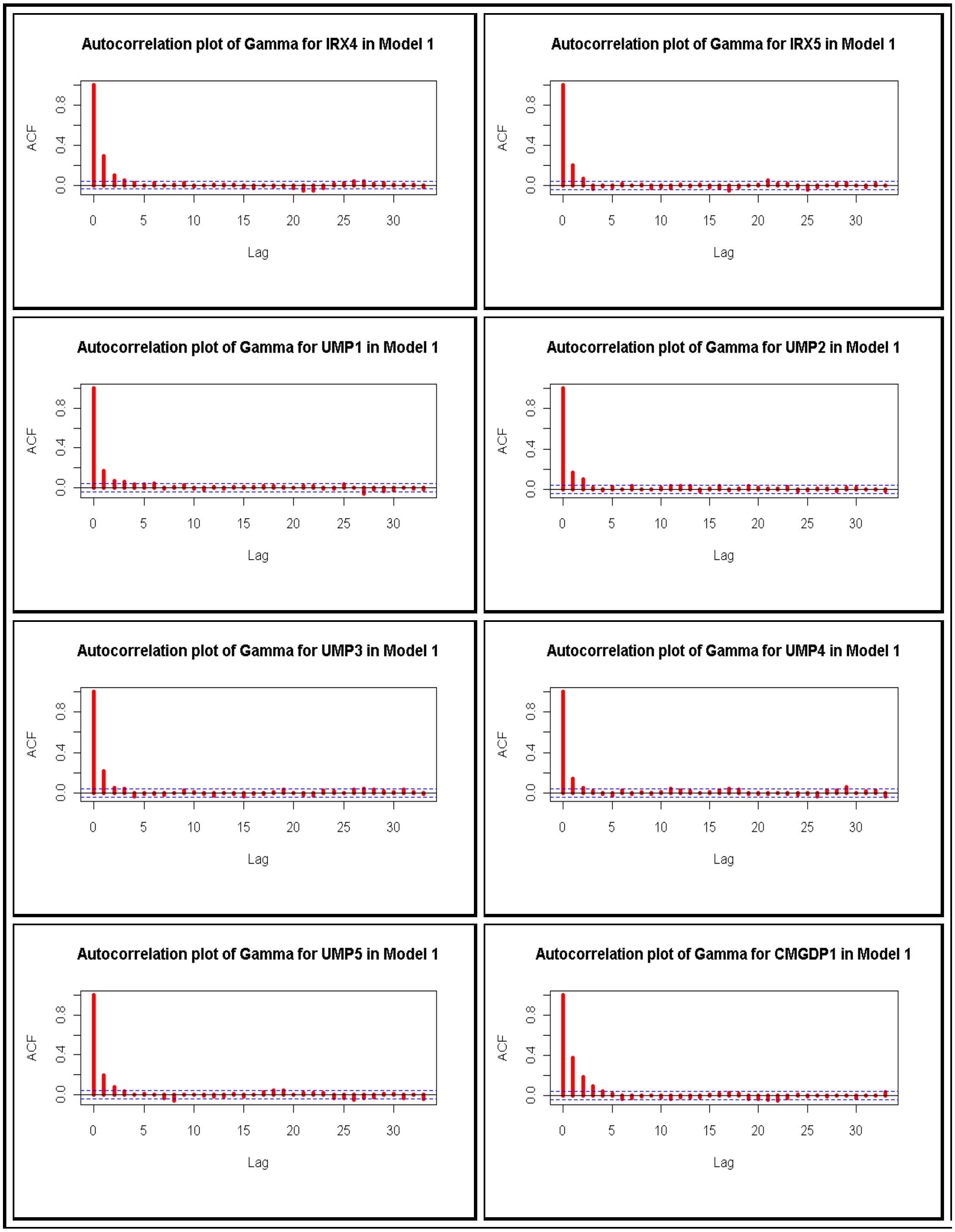
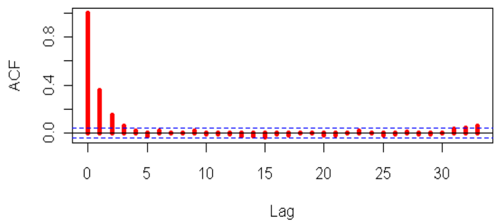
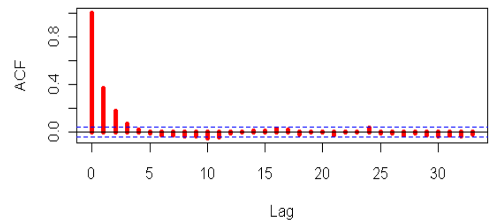

Autocorrelation plot of Gamma for IRX4 in Model 1
Autocorrelation plot of Gamma for IRX5 in Model 1
Autocorrelation plot of Gamma for UMP1 in Model 1
Autocorrelation plot of Gamma for UMP2 in Model 1
Autocorrelation plot of Gamma for UMP3 in Model 1
Autocorrelation plot of Gamma for UMP4 in Model 1
Autocorrelation plot of Gamma for UMP5 in Model 1
Autocorrelation plot of Gamma for CMGDP1 in Model 1
ACF
0.0
0.4
0.8
Lag
0
5
10
15
20
25
30

Table E continued:

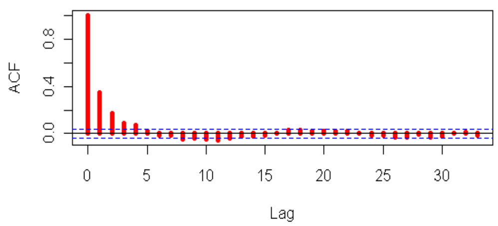
Autocorrelation plot of Gamma for CMGDP2 in Model 1



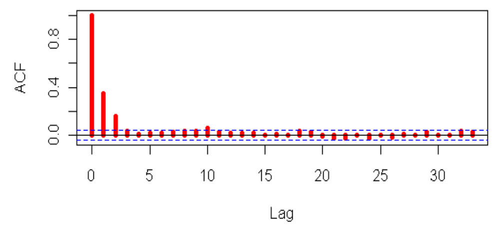
Autocorrelation plot of Gamma for CMGDP3 in Model 1



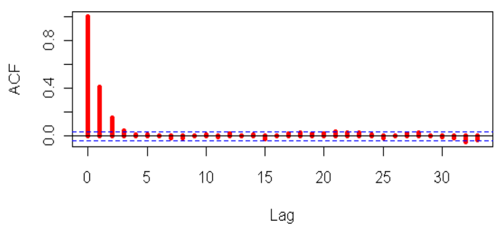
Autocorrelation plot of Gamma for CMGDP4 in Model 1



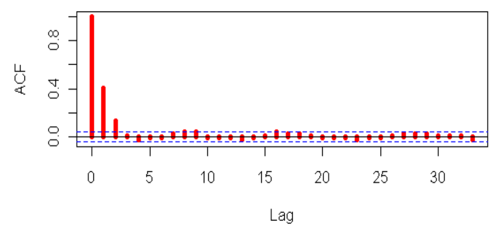
Autocorrelation plot of Gamma for CMGDP5 in Model 1



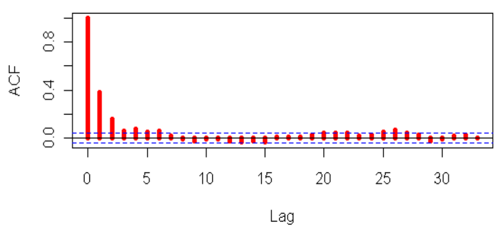
Autocorrelation plot of Gamma for HUR1 in Model 1



Autocorrelation plot of Gamma for HUR2 in Model 1



Autocorrelation plot of Gamma for HUR3 in Model 1



Autocorrelation plot of Gamma for HUR4 in Model 1

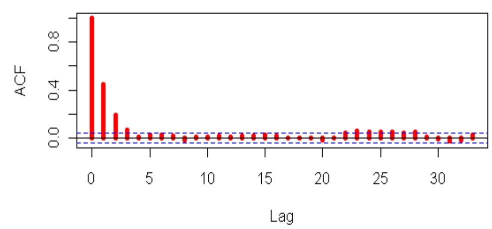


Table E continued:

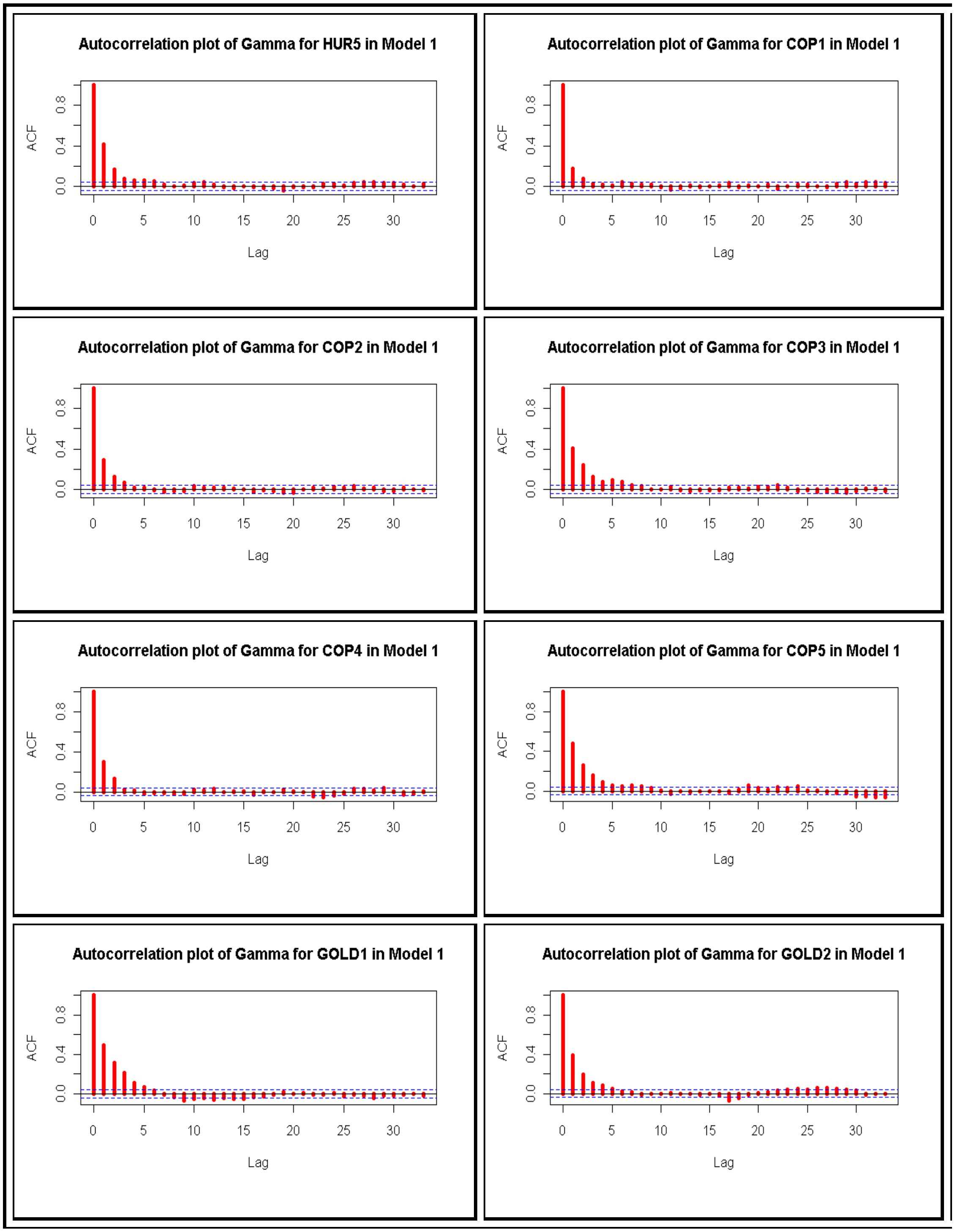


Table E continued:

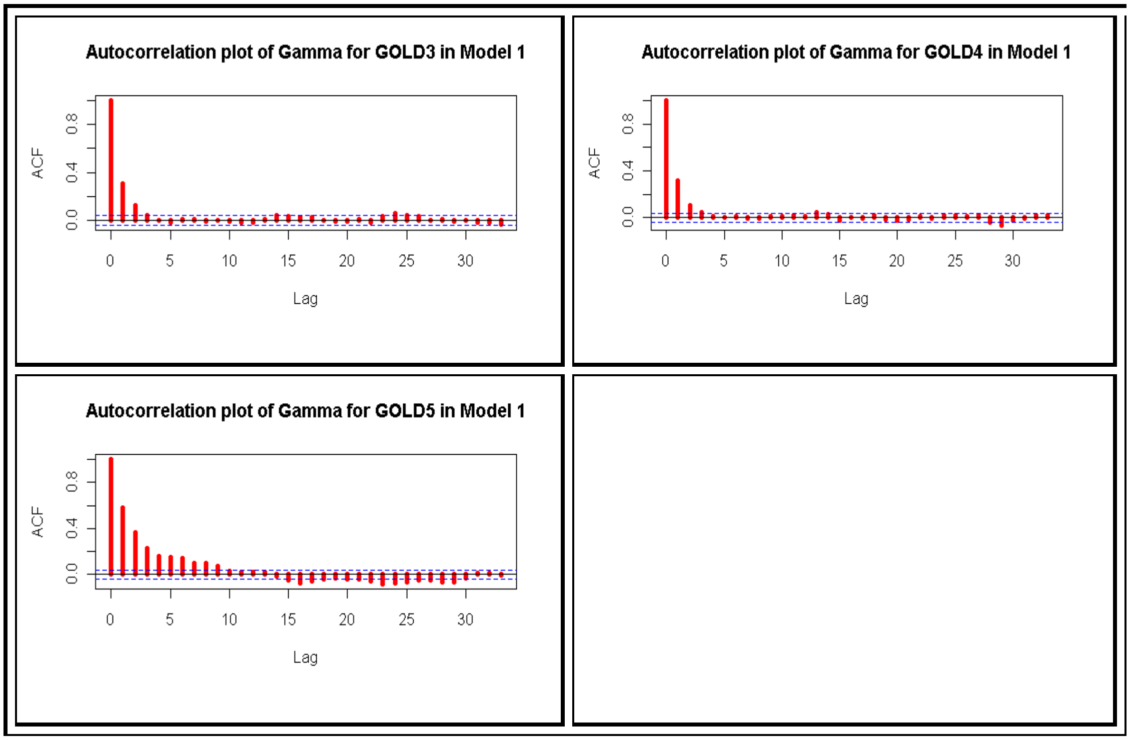


Table F: History plot for the β's in model 2

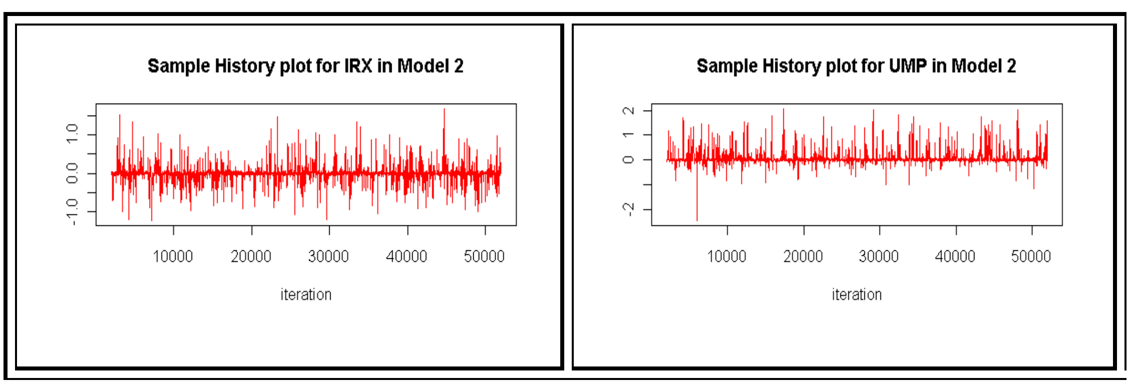
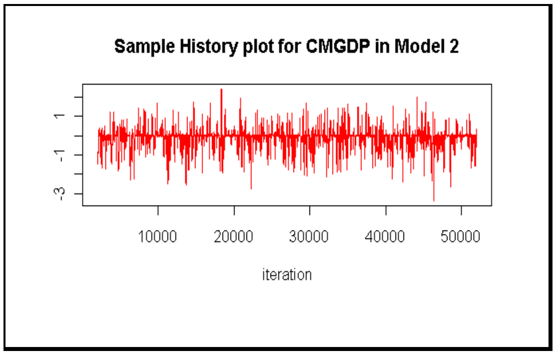
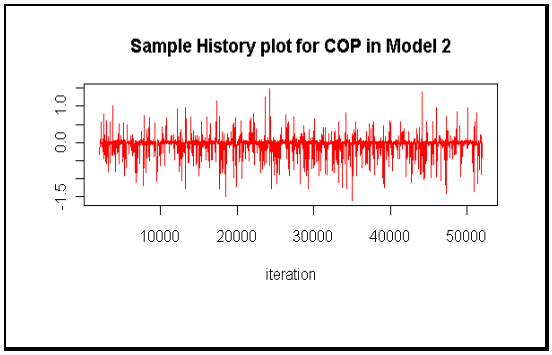
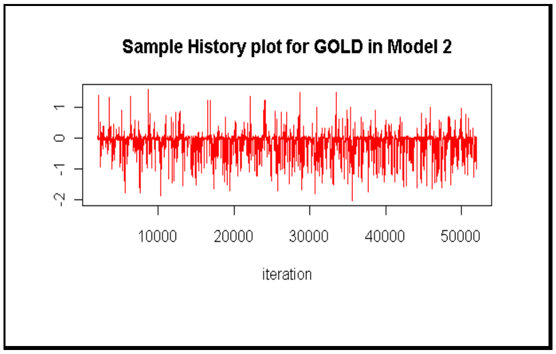
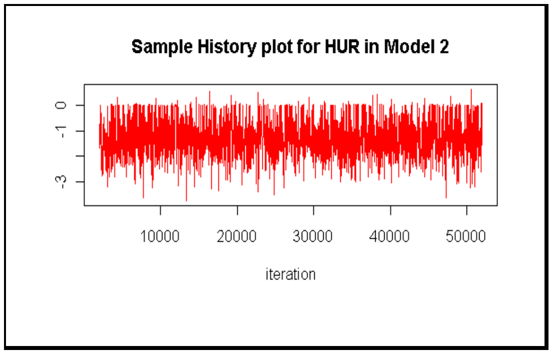


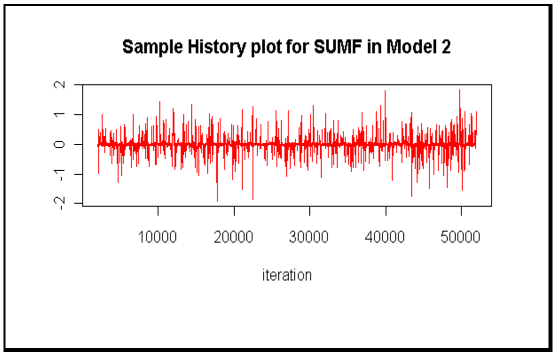
Table F continued:

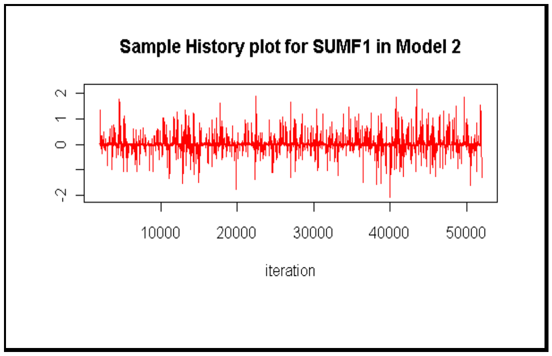


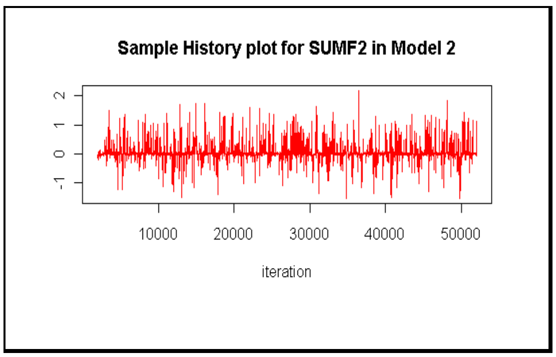












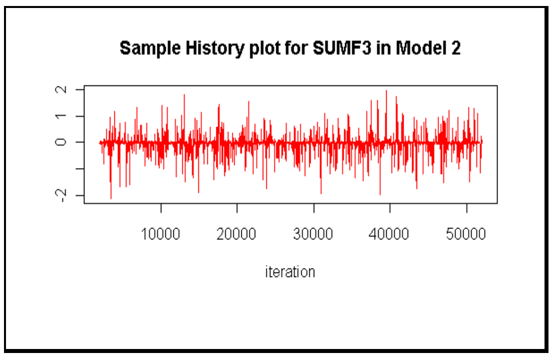
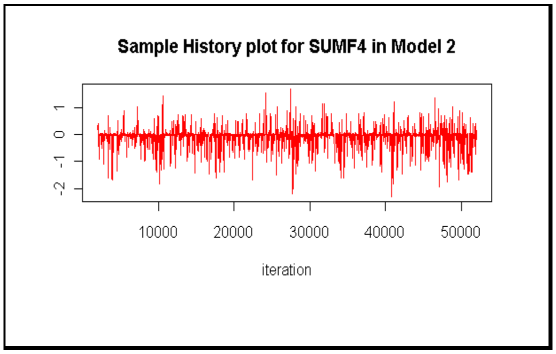
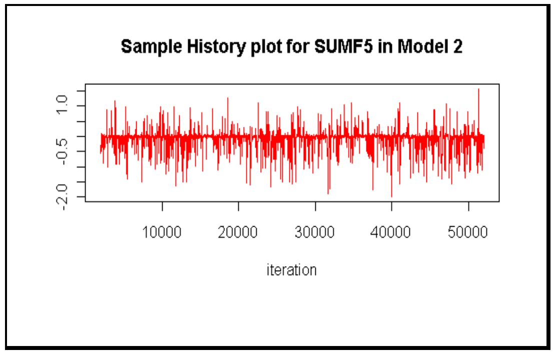
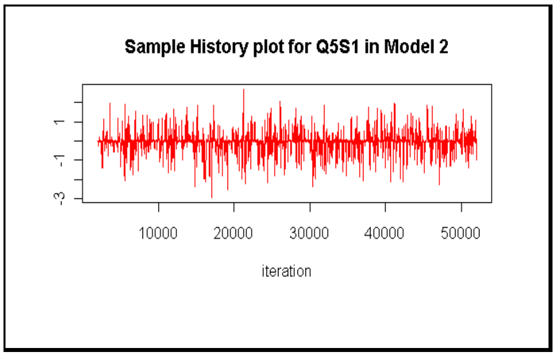
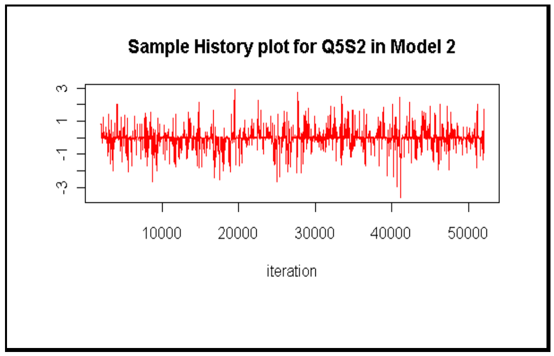


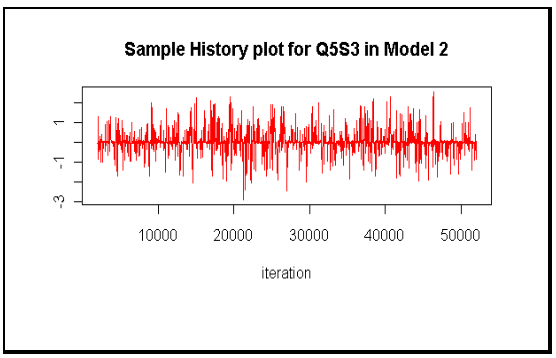
Table F continued:

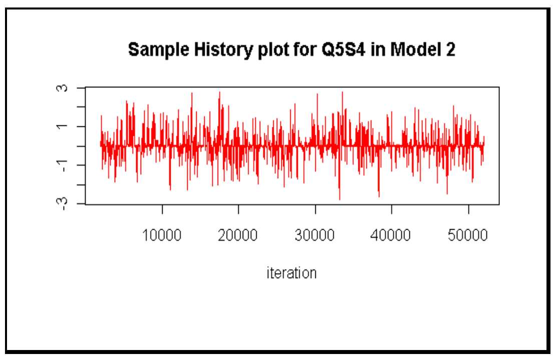


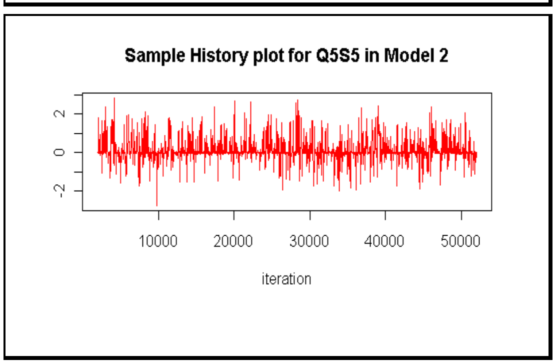












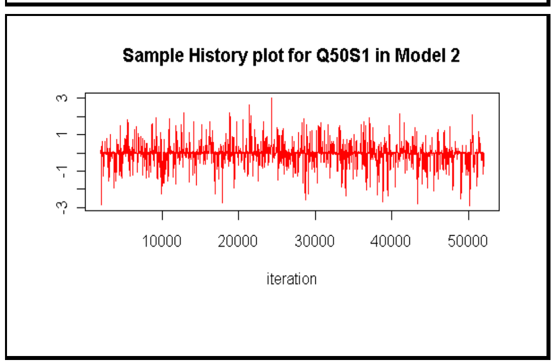
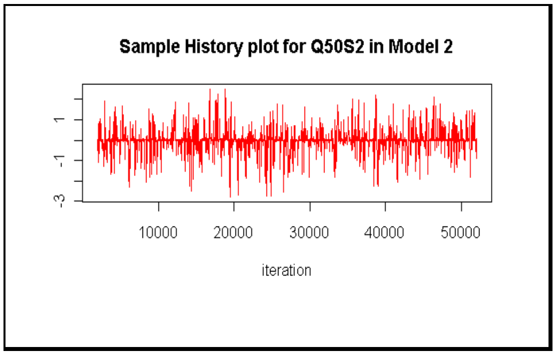
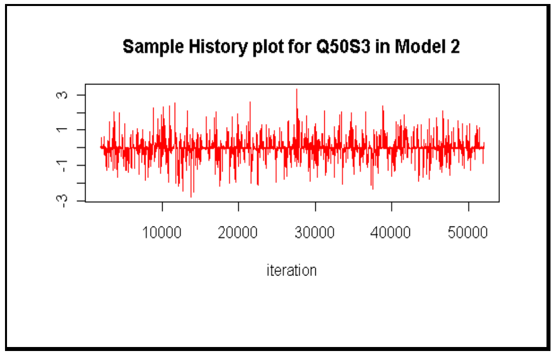
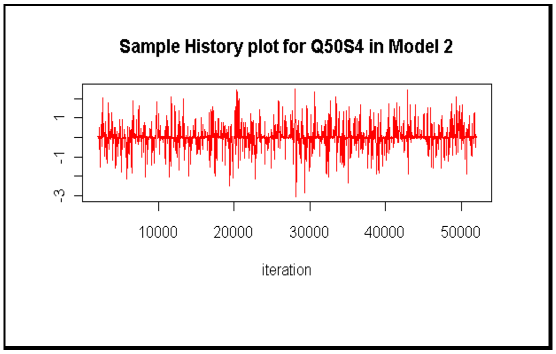
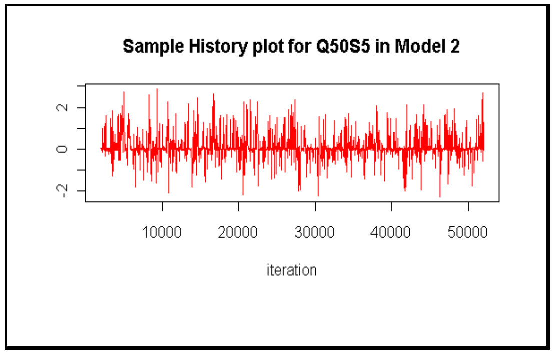


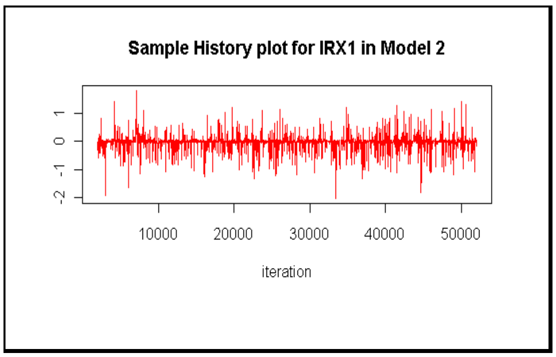
Table F continued:

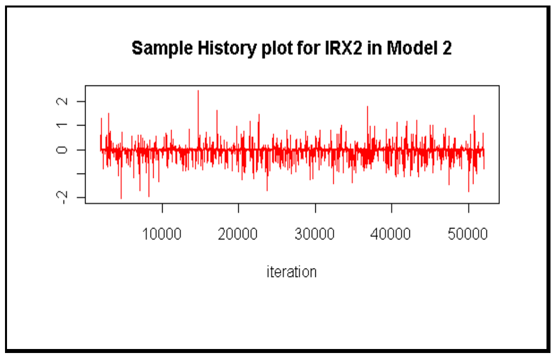


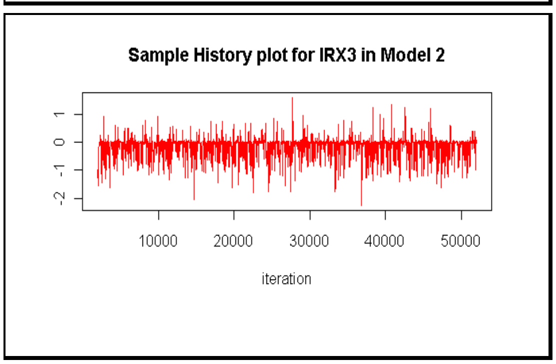












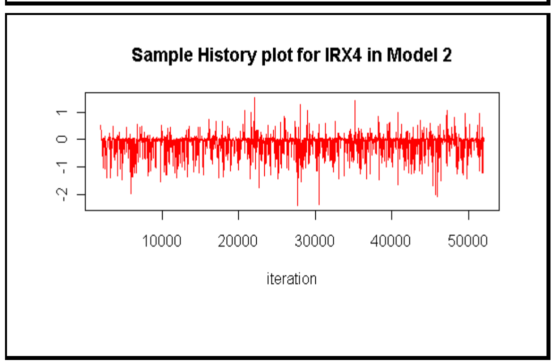


Table F continued:

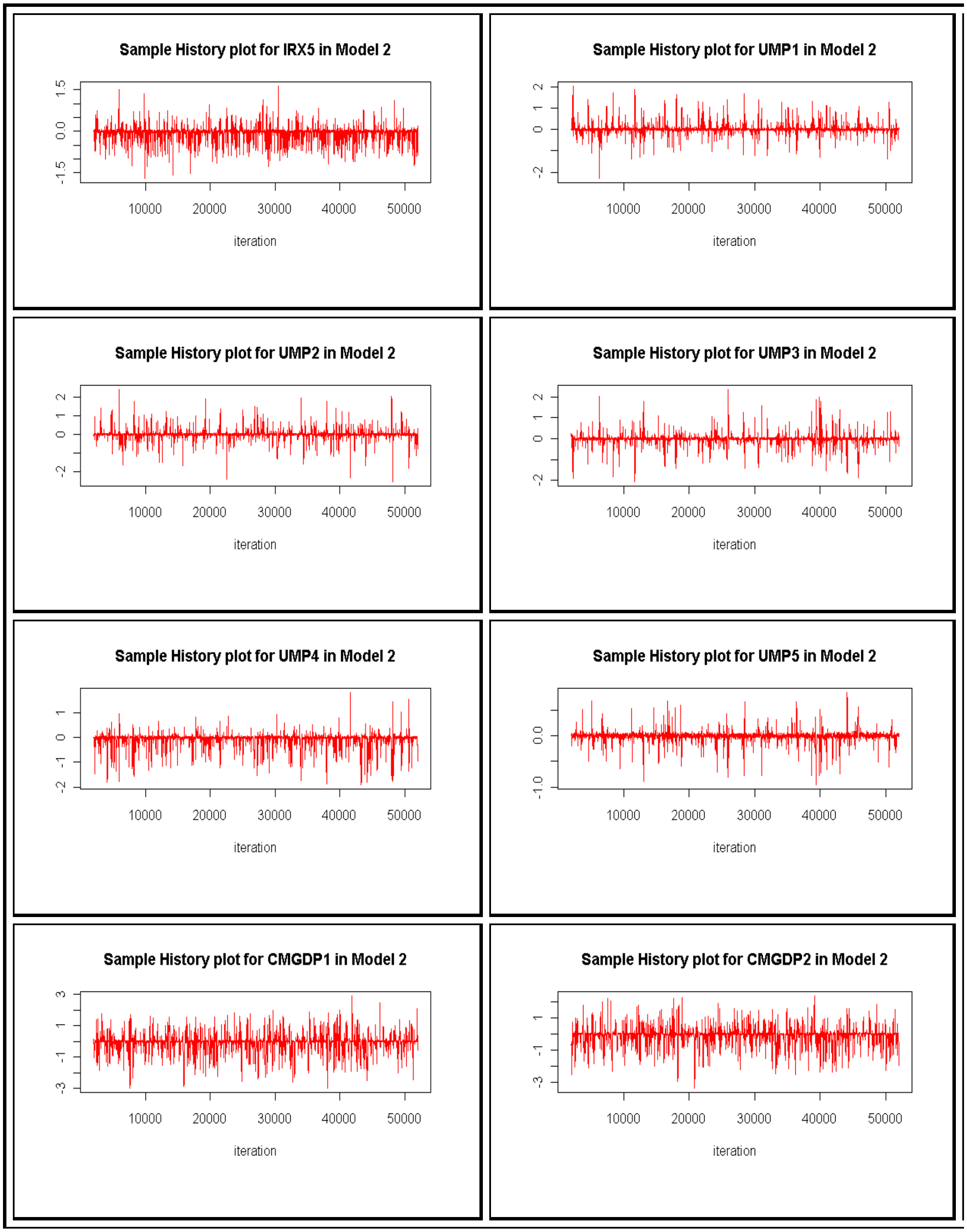


Table F continued:

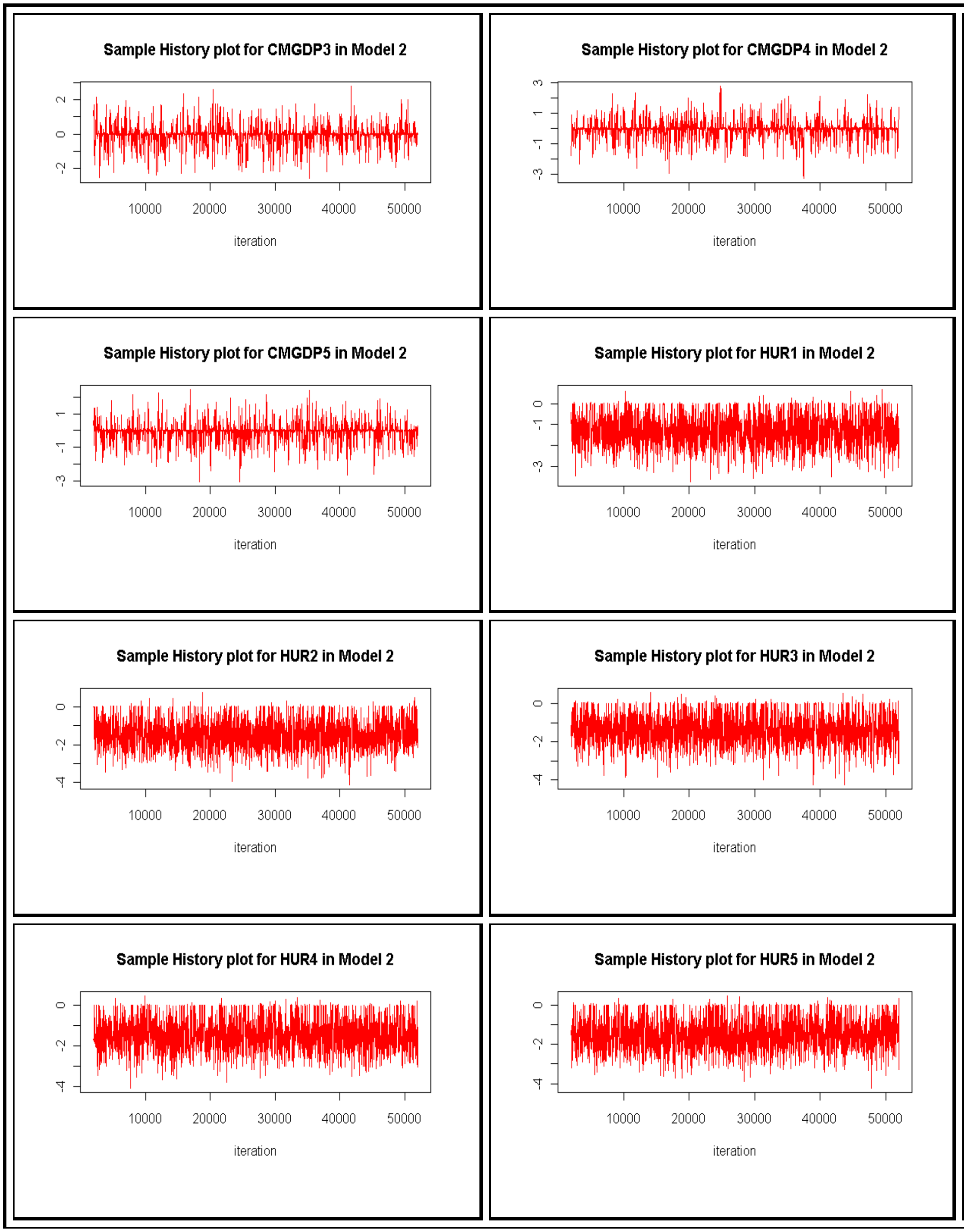


Table F continued:

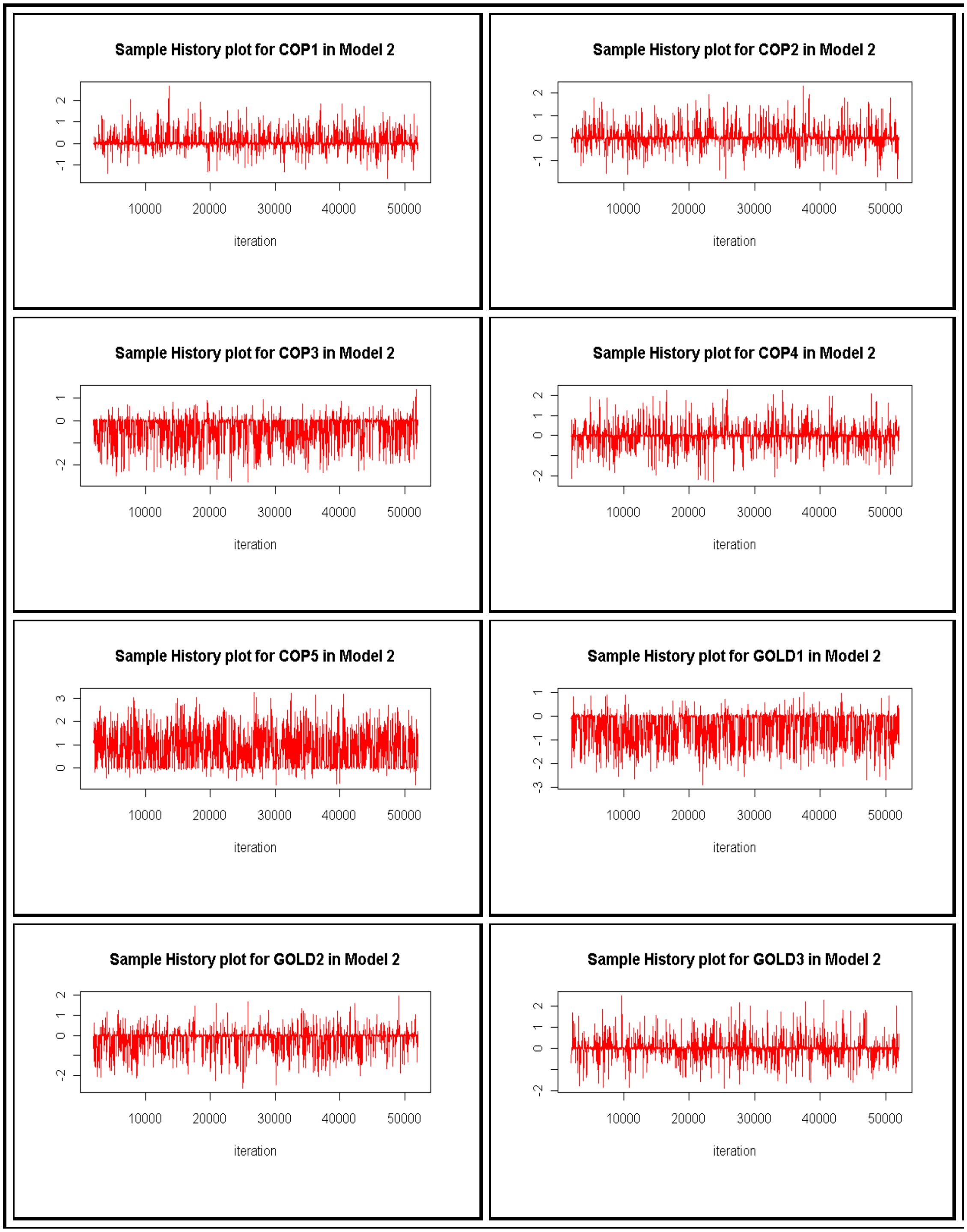
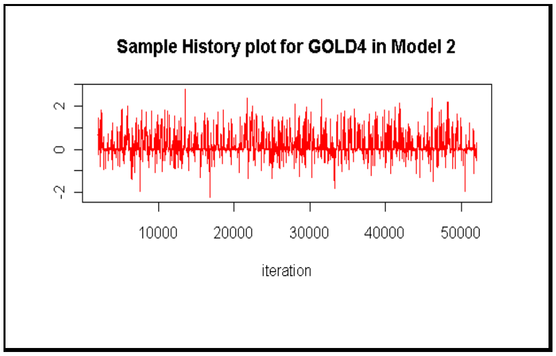


Table F continued:



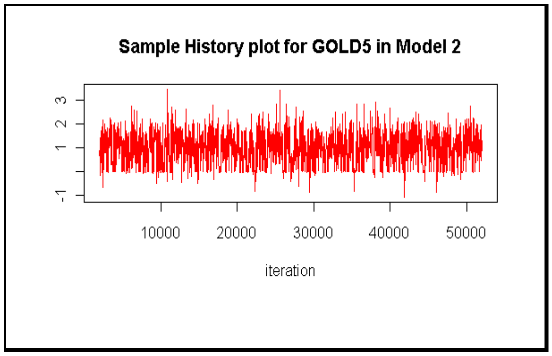
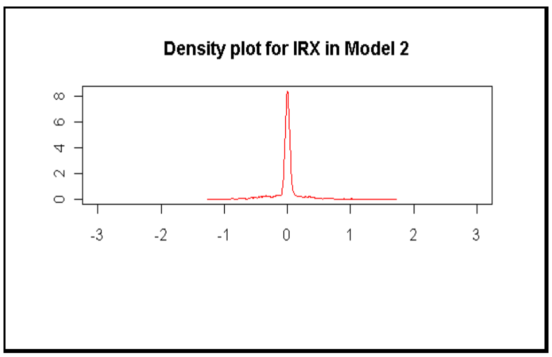
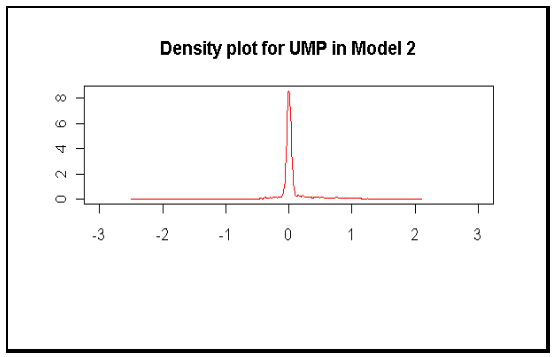
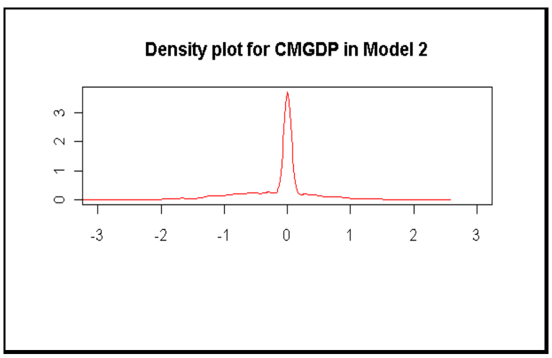


Table G: Density plot for the β's in model 2







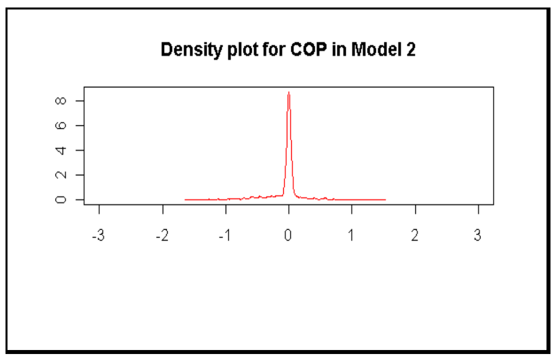


Table G continued:

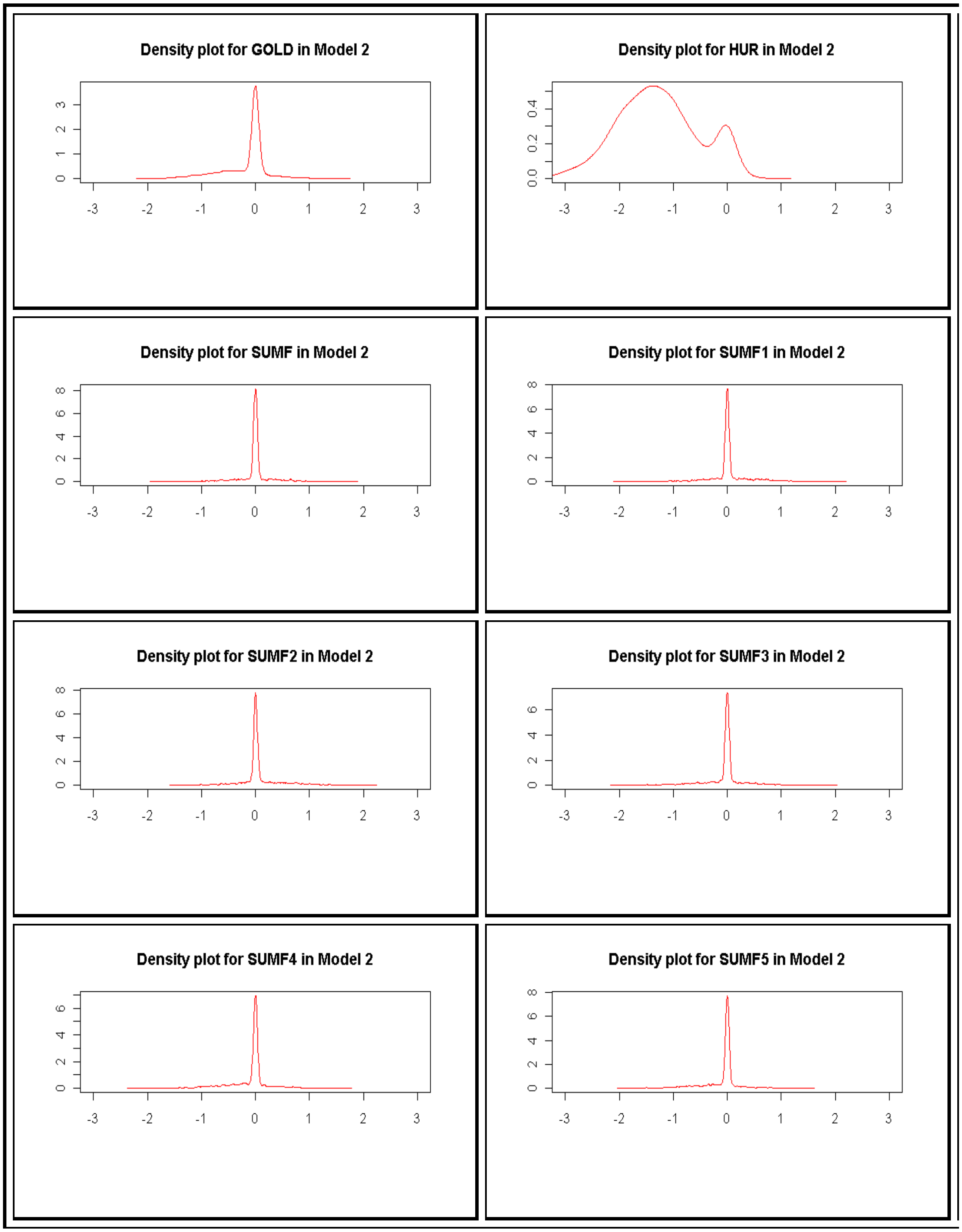
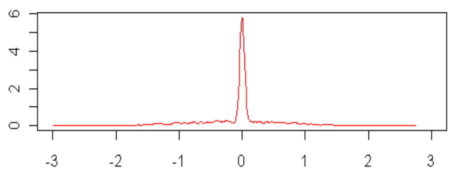
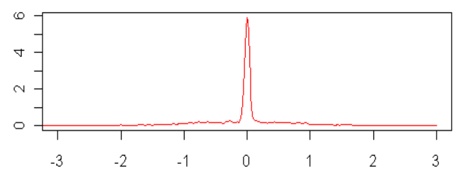


Table G continued:

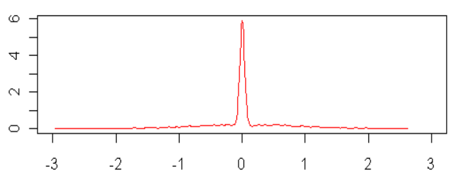
Density plot for Q5S1 in Model 2



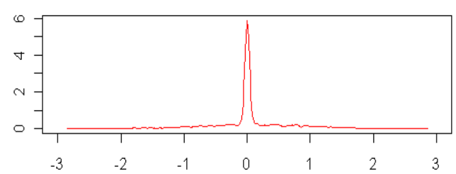
Density plot for Q5S2 in Model 2



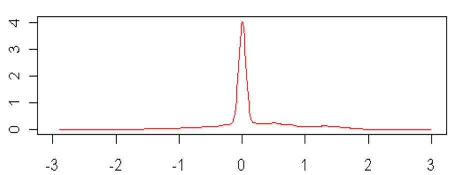
Density plot for Q5S3 in Model 2



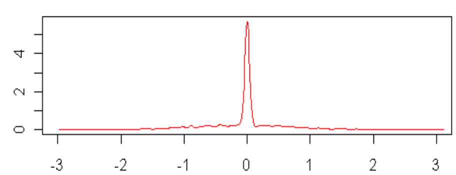
Density plot for Q5S4 in Model 2



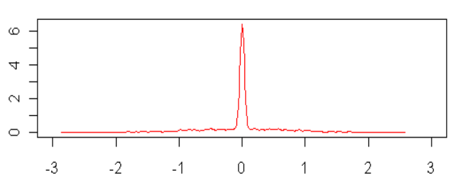
Density plot for Q5S5 in Model 2



Density plot for Q50S1 in Model 2



Density plot for Q50S2 in Model 2



Density plot for Q50S3 in Model 2

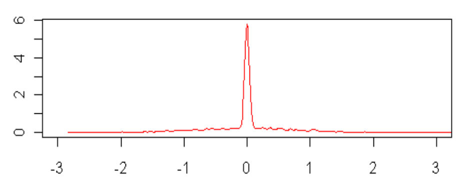


Table G continued:

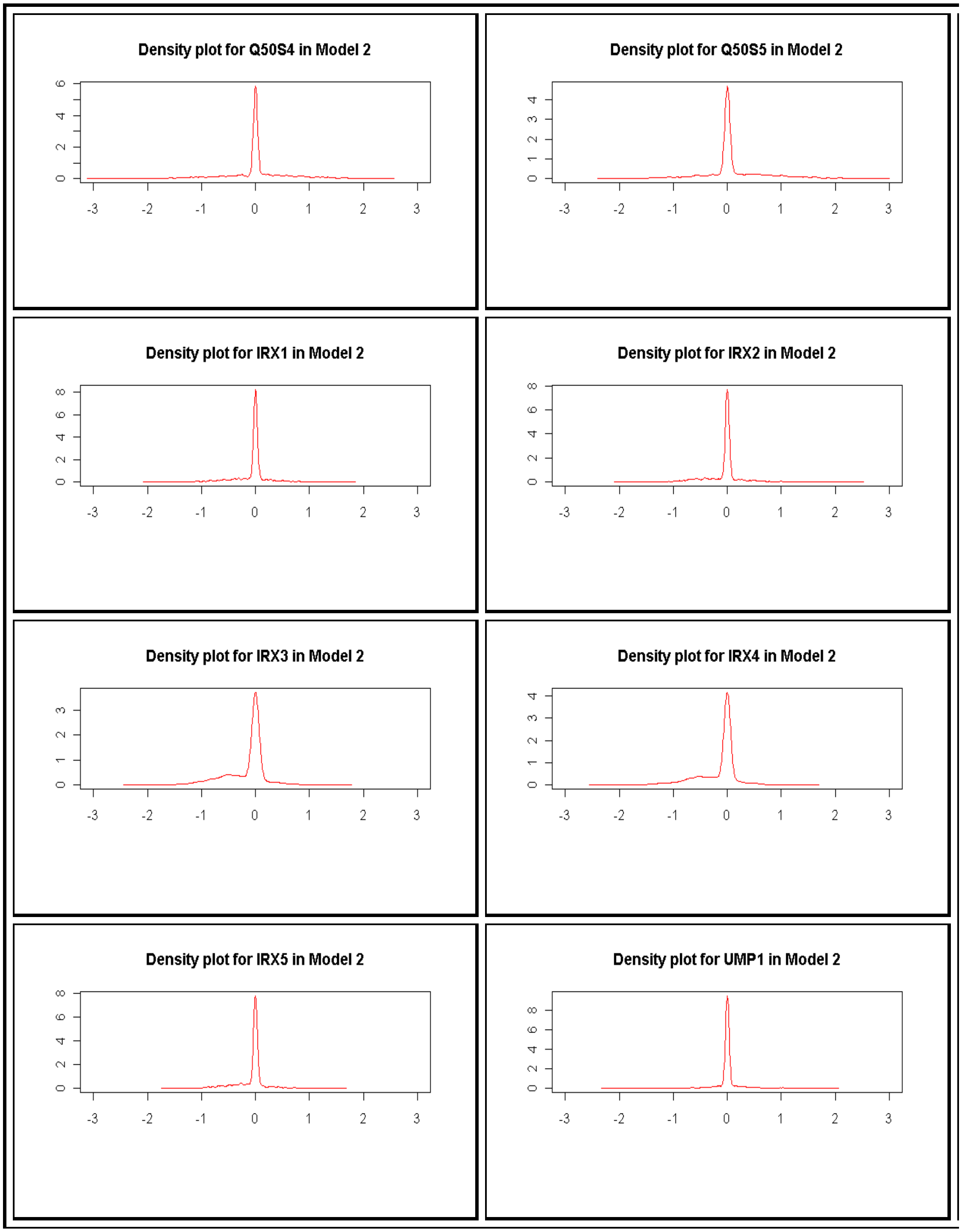


Table G continued:

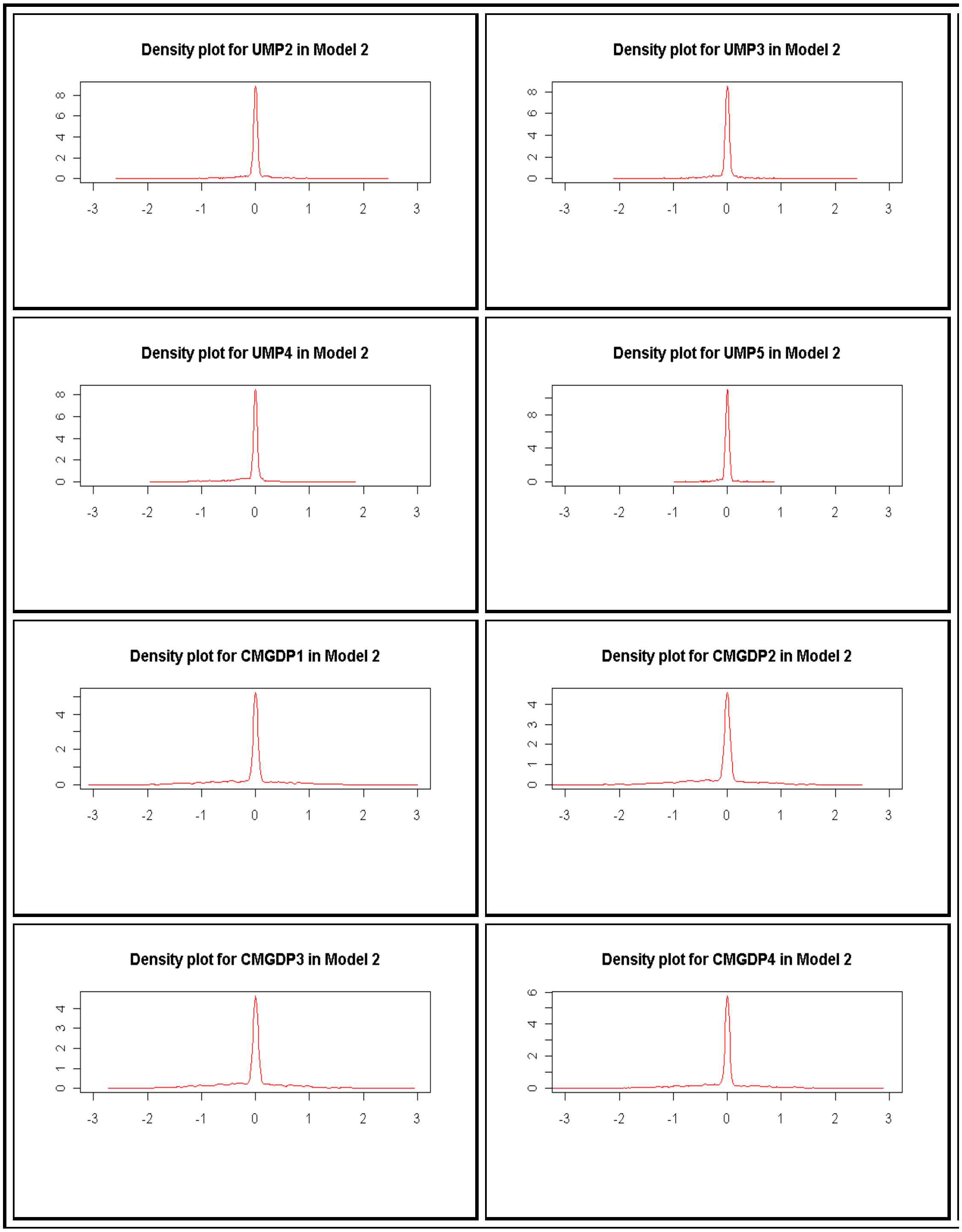


Table G continued:

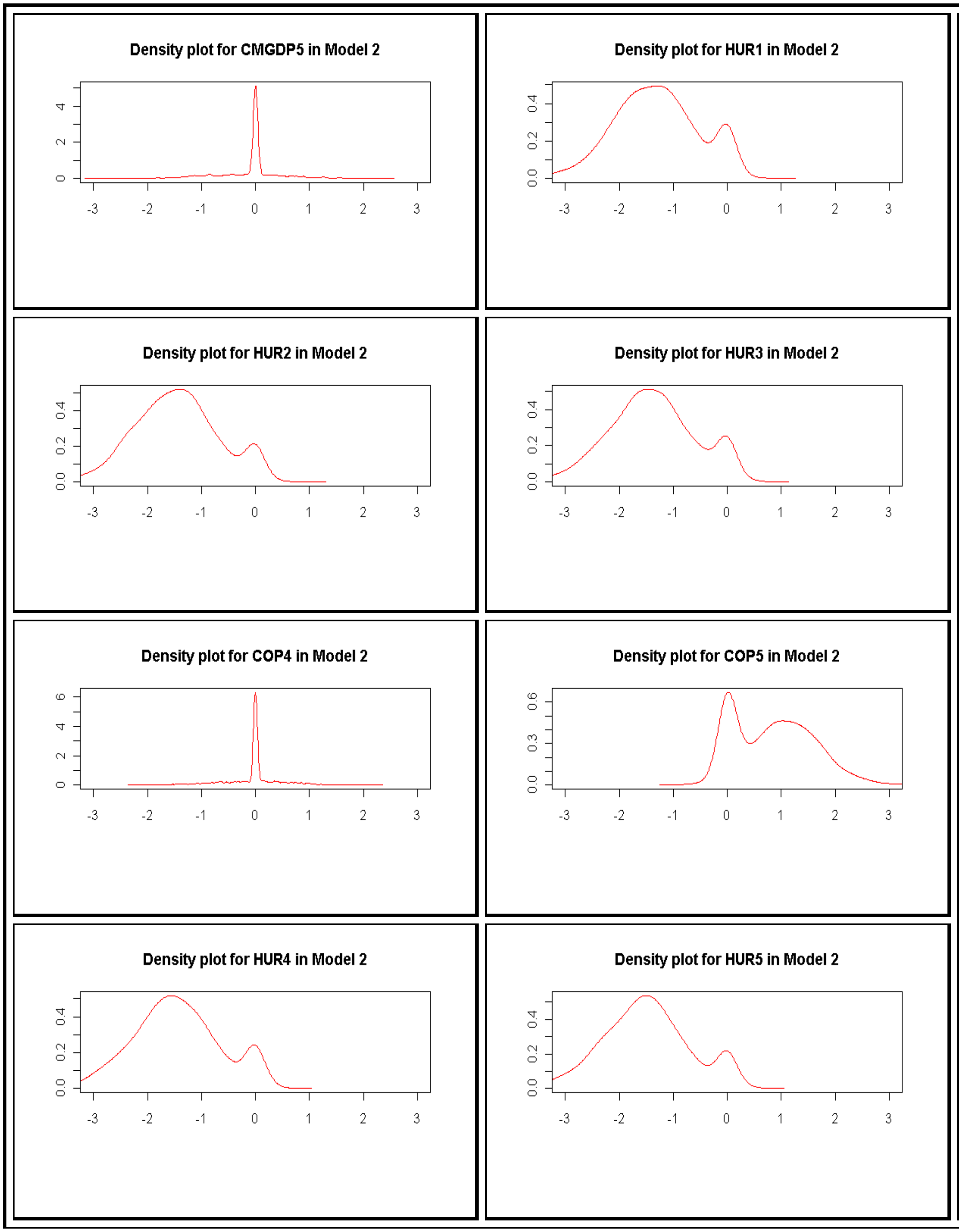


Table G continued:

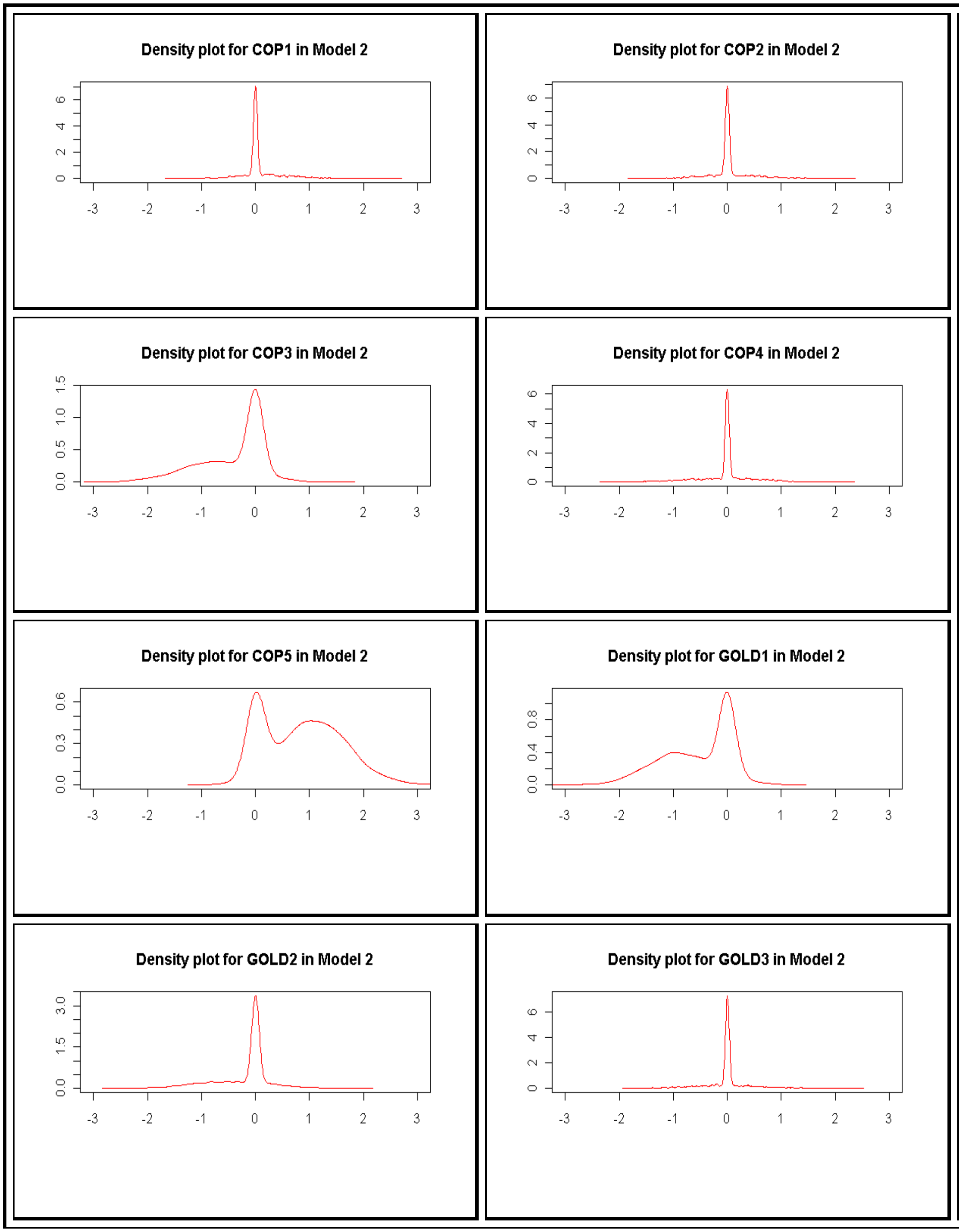


Table G continued:

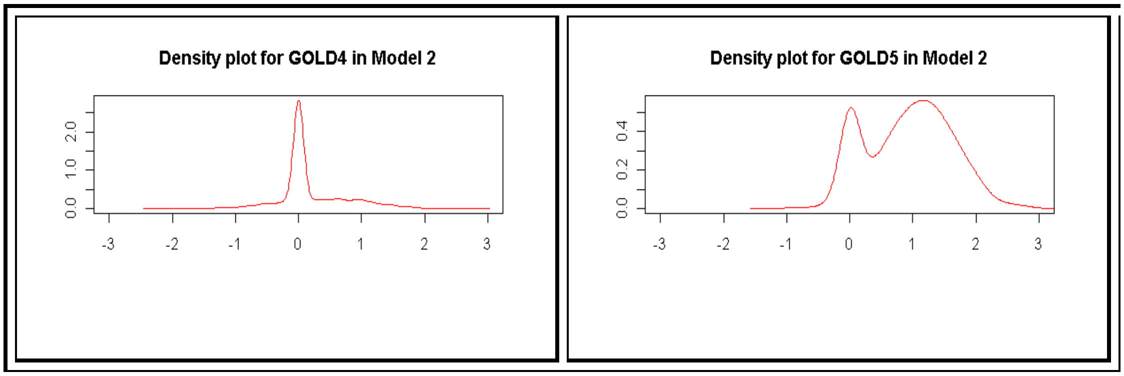


Table H: Autocorrelation plot for the γ's in model 1

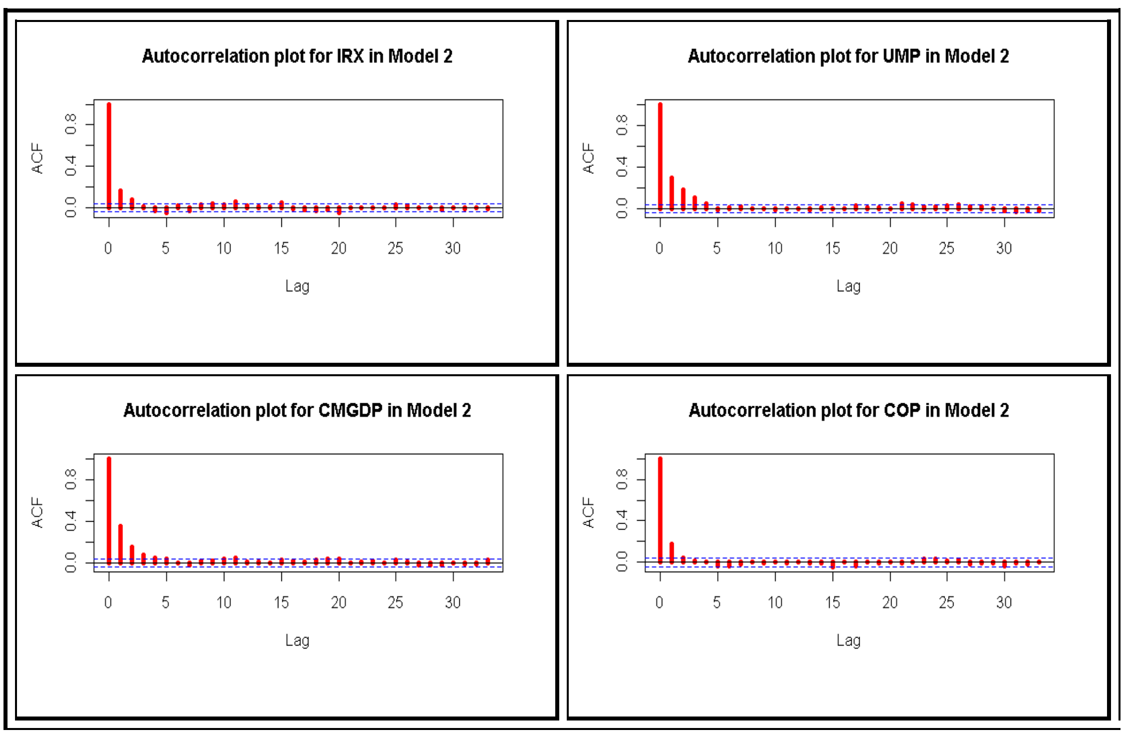


Table H continued:

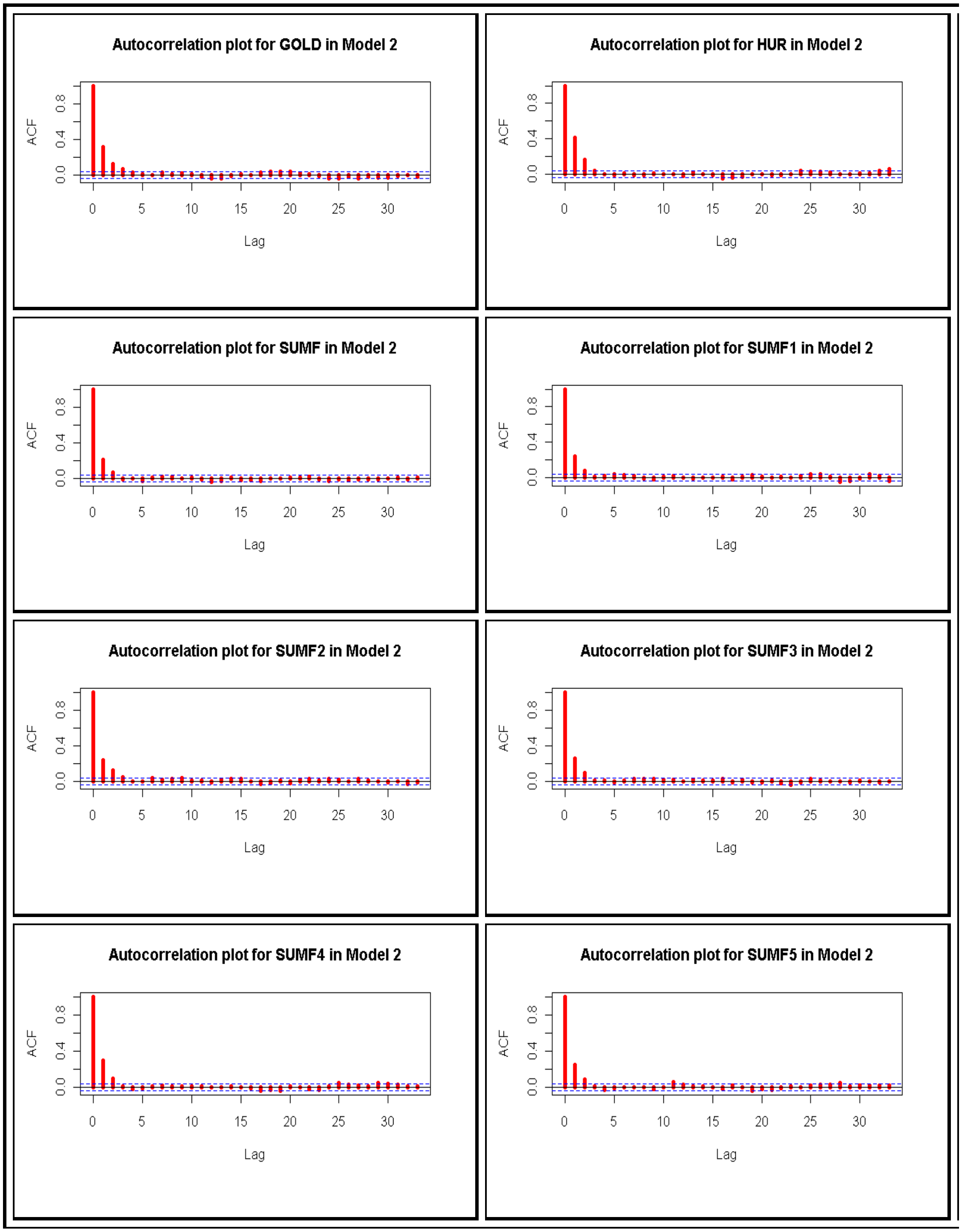
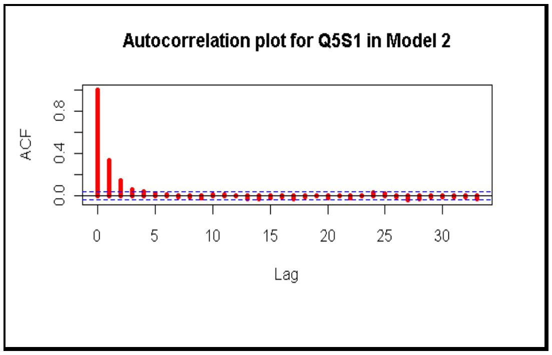
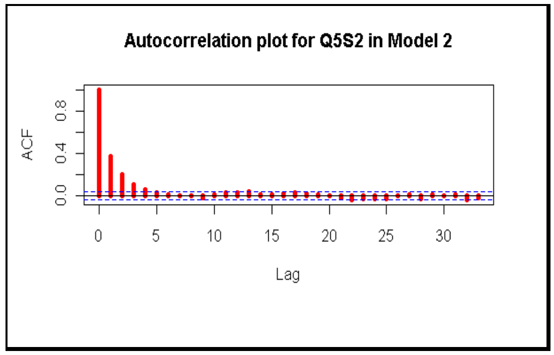
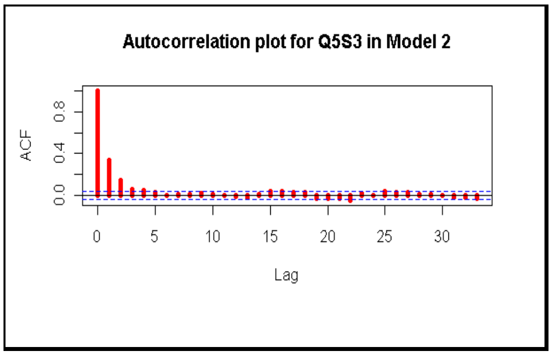
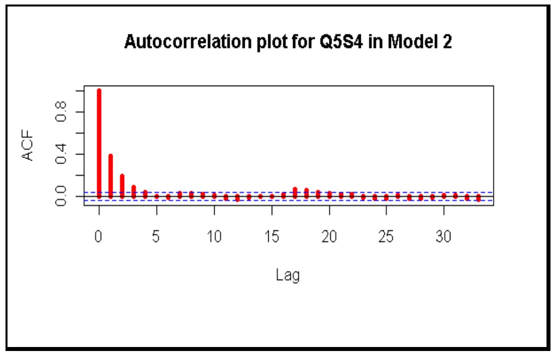


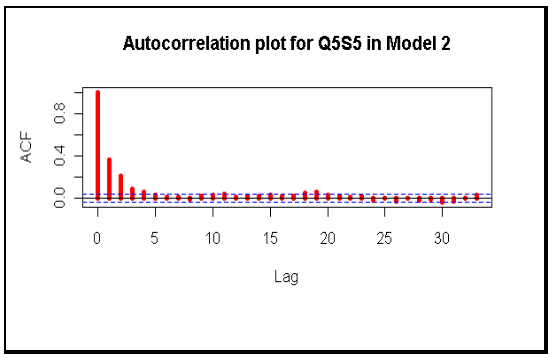
Table G continued:

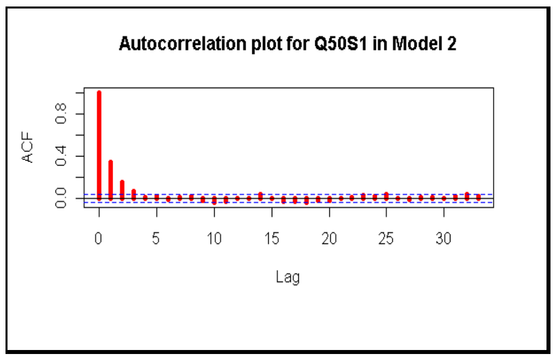


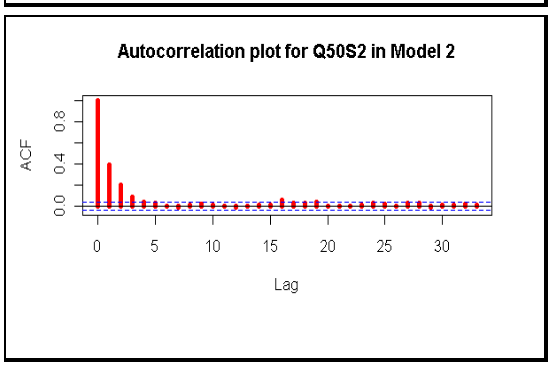












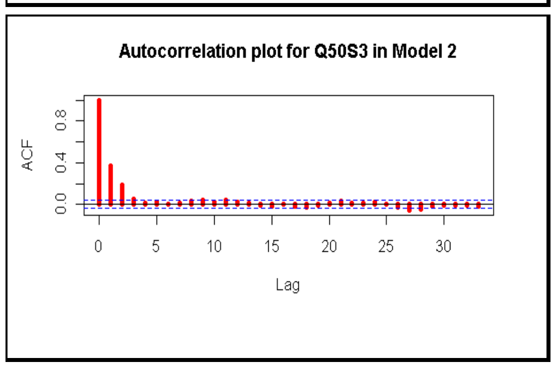
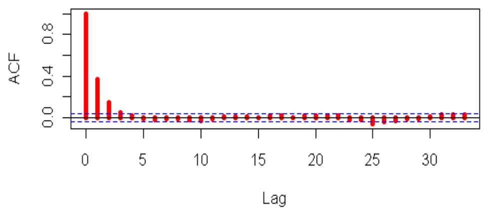
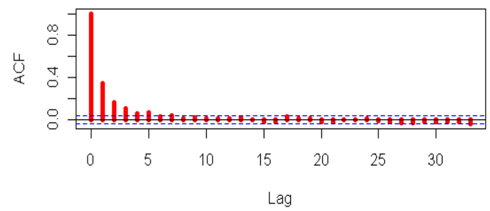


Table G continued:

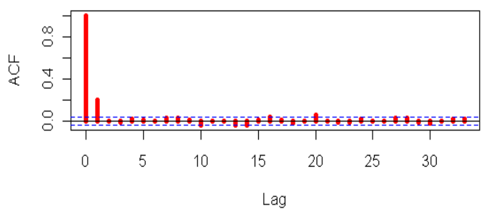
Autocorrelation plot for Q50S4 in Model 2



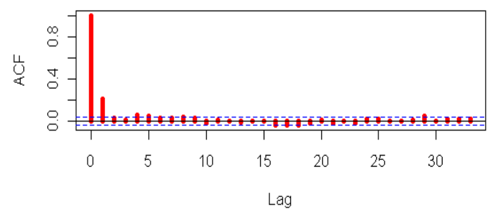
Autocorrelation plot for Q50S5 in Model 2



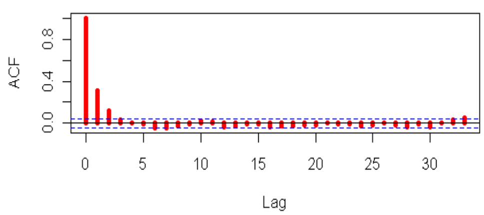
Autocorrelation plot for IRX1 in Model 2



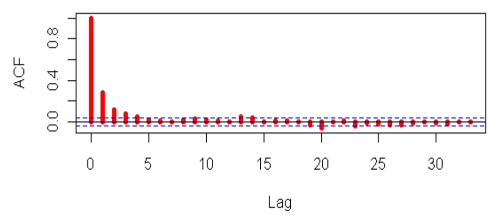
Autocorrelation plot for IRX2 in Model 2



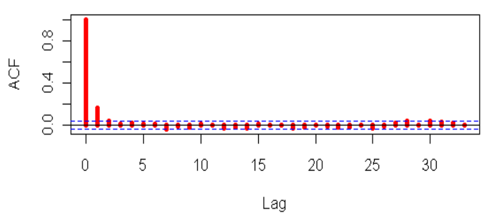
Autocorrelation plot for IRX3 in Model 2



Autocorrelation plot for IRX4 in Model 2



Autocorrelation plot for IRX5 in Model 2



Autocorrelation plot for UMP1 in Model 2

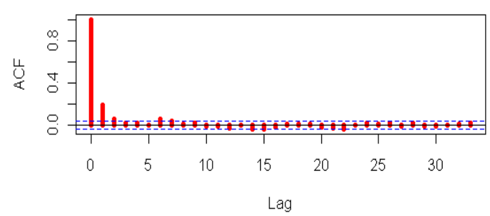
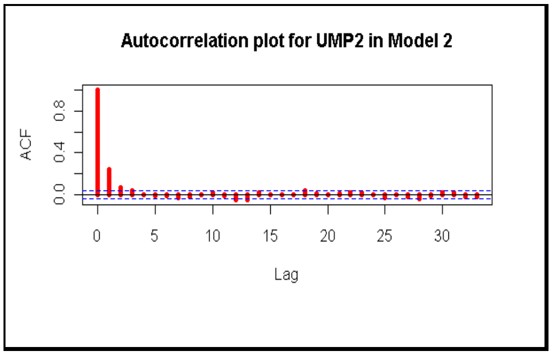
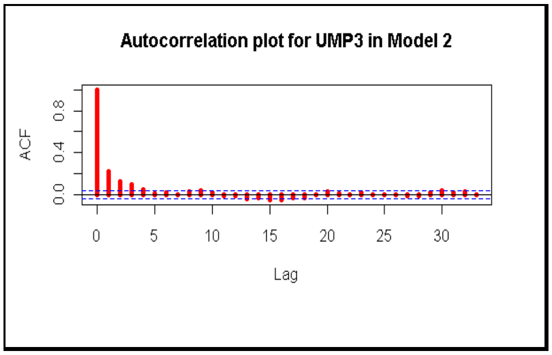
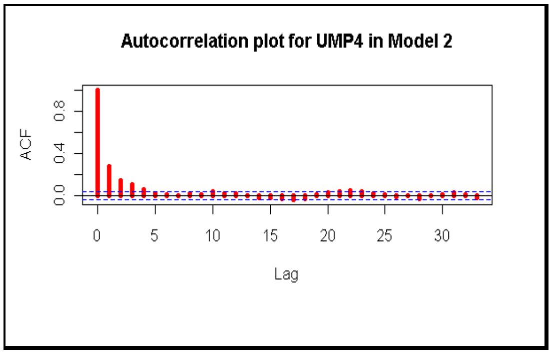
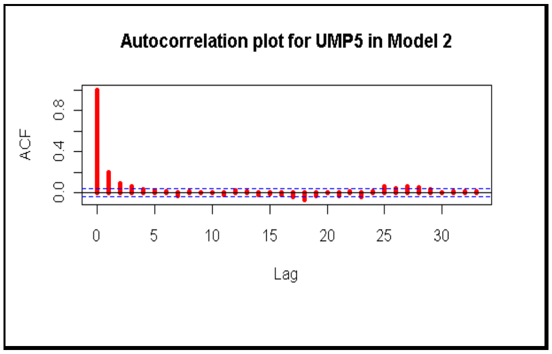


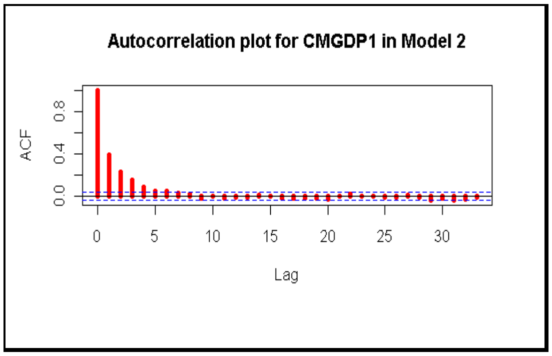
Table G continued:

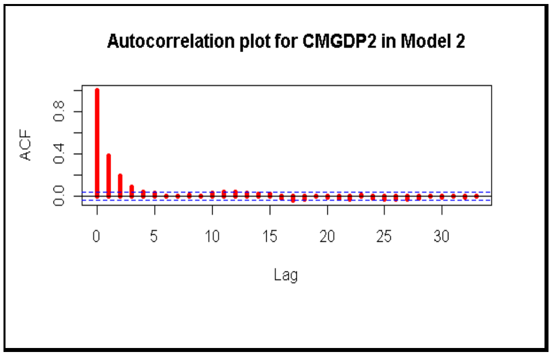


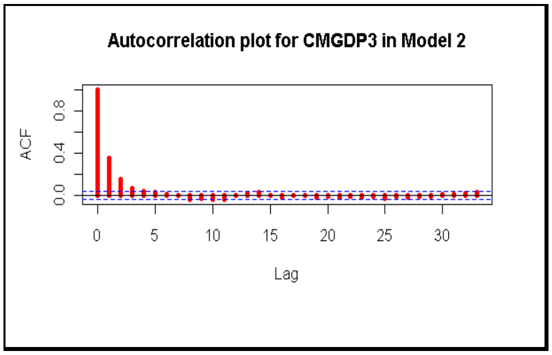












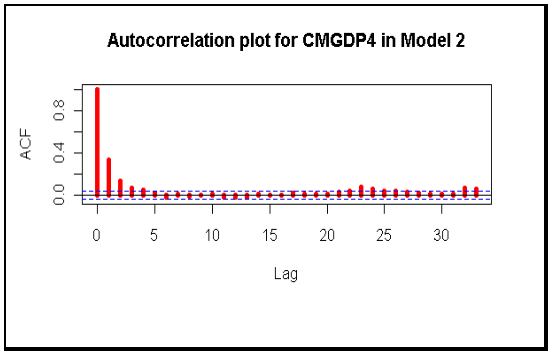
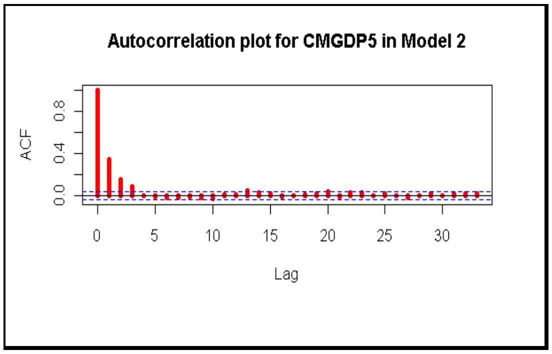
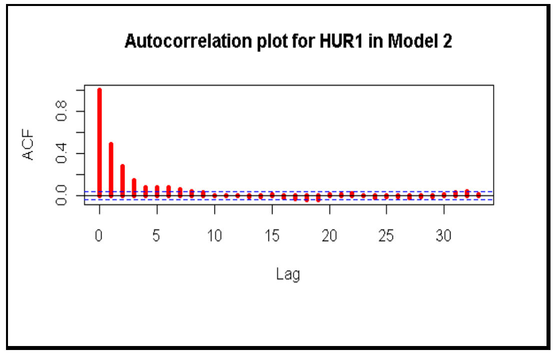
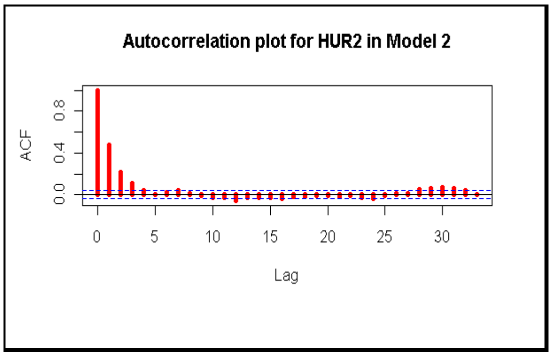
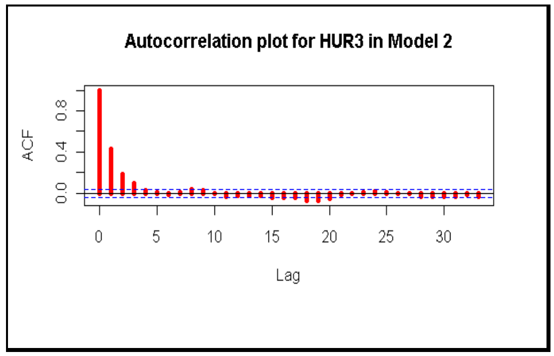


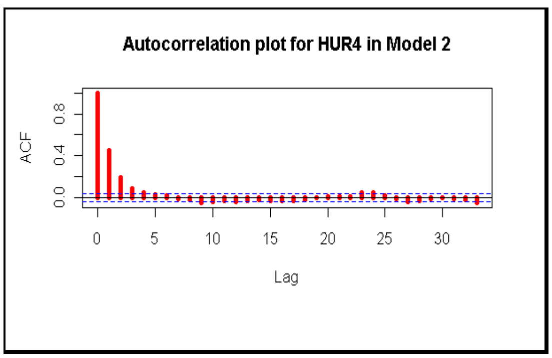
Table G continued:

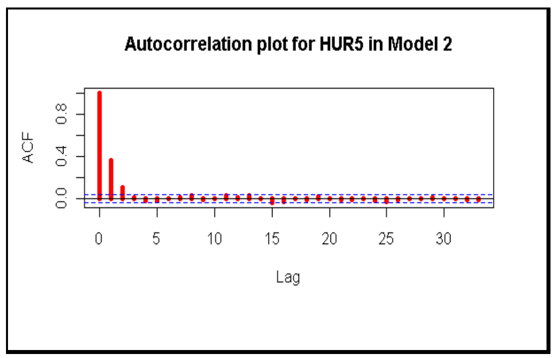


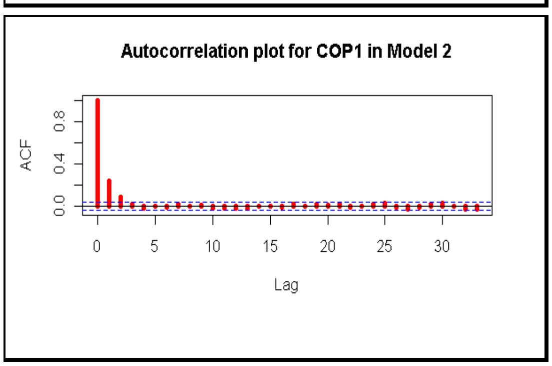












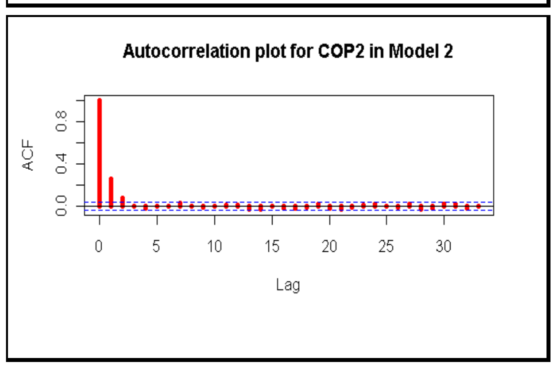
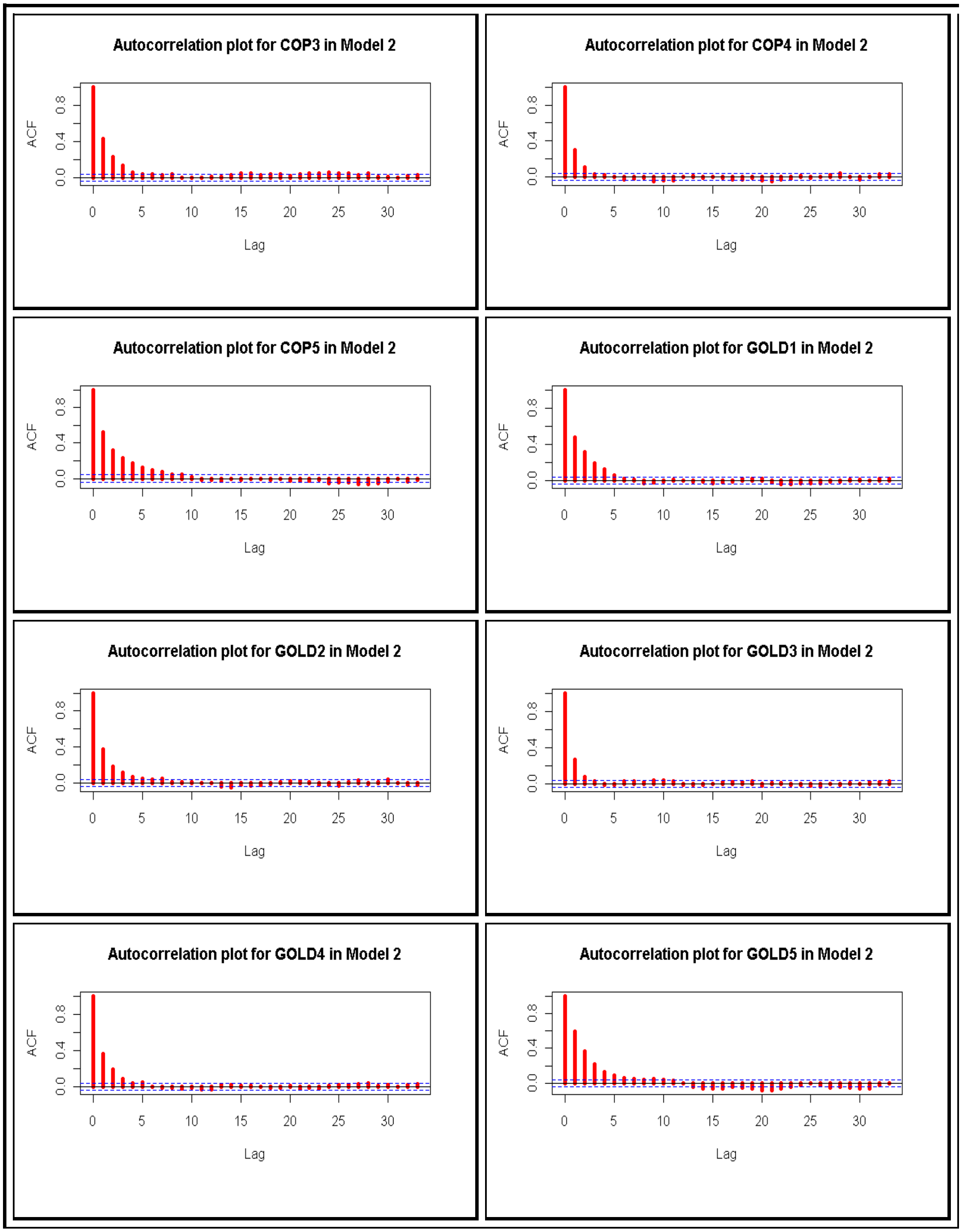


Table G continued:



Bibliography

[1]. Financial crisis in 1974

http://en.wikipedia.org/wiki/1973%E2%80%931974_stock_market_crash

[2]. Black Monday http://recessionhistory.info/black-monday/

http://en.wikipedia.org/wiki/Black_Monday_(1987)

[3]. Early 2000's recession

http://en.wikipedia.org/wiki/Early_2000s_recession#United_States

[4]. Subprime mortgage crisis http://en.wikipedia.org/wiki/Subprime_mortgage_crisis

[5]. Unemployment rate

http://research.stlouisfed.org/fred2/series/UNRATE/downloaddata

[6]. IRX index monthly http://finance.yahoo.com/q?s=^irx

[7]. Price of West Texas Intermediate Crud http://www.economagic.com/em-cgi/data.exe/var/west-texas-crude-long

[8]. Gold price index, monthly http://www.research.gold.org/prices/

[9]. Total credit market debt as percentage of GDP http://www.economagic.com/em-cgi/data.exe/var/togdp-totalcreditdebt

[10]. Business Cycle http://en.wikipedia.org/wiki/Business_cycle

[11]. Historical Major Hurricane http://www.nhc.noaa.gov/HAW2/english/history.shtml

[12]. George, E. I., McCulloch R. E., *Variable Selection via Gibbs Sampling.* Journal of American Statistical Association, 1993. **88:** p. 881-889

[13]. Ntzoufras, I., Forster, J. J. and Dellaportas, P., *Stochastic search variable selection for log-linear models.* Journal of Statistical Computation and Simulation, 2000. **68:** p. 23-27

[14]. Swartz, M. D., Kimmel, M., Mueller, P., and Amos, C. I., *Stochastic search gene suggestion: A Bayesian hierarchical model for gene mapping.* Biometrics, 2006. **62:** p. 495-503

[15]. Swartz M. D., Yu, R. K. and Shete, S., *Finding factors influencing risk: Comparing Bayesian stochastic search and standard variable selection methods applied to logistic regression models of cases and controls.* Statistics in Medicine, 2008. **27:** p. 6158-6174

[16]. Barbieri, M. M. and Berger, J. O., *Optimal predictive model selection.* Annals of Statistics, 2004. **32:** p. 870-897

[17]. R-project http://www.r-project.org/ BRugs

Vita

Chi-san Ho was born in Taipei, Taiwan on 7 September 1982, the daughter of Shan-ping Ho and Hsin-ping Hsu, and the niece of Shan-yuan Ho.

She received her Bachelor of Arts degree in Business Administration from National Taiwan University (NTU). She developed her interest in statistics from various management and economics courses. She was also selected to be an exchange student to spend her senior year in San Francisco, USA.

After obtaining her Bachelor's degree she started working in the Taiwan Economic Data Center (TEDC) as a research assistant for two years. Working with Prof. M. Liang, the CEO of TEDC, she gained a lot of experience in data processing and human resource management.

In the summer of 2007, she started graduate school in Statistics at the University of Texas at Austin, where she studied Bayesian statistics. She will join the Information, Risk and Operations Management Department (IROM) at McCombs Business School in fall 2009 for doctorate study in Risk Management and Statistics.

Permanent address: 3F, No. 20, Lane 12, Wenzhou St.

Taipei, Taiwan, 106.

This report was typed by Chi-san Ho.